

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2003, 04:39:33 ; Search time 188 Seconds  
(without alignments)  
8439.558 Million cell updates/sec

Title: US-10-010-160-1

Perfect score: 622

Sequence: 1 atgtctgatgatccagctaa.....aatatctgcagtcgacagtc 622

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 1708419 seqs, 1275431651 residues 3416838

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09C\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10B\_NEW\_PUB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	622	100.0	622	US-10-010-160-1	Sequence 1, Appl1
2	54.8	8.8	9052	US-10-311-455-118	Sequence 118, Appl
3	53	8.5	945	US-09-895-913A-45	Sequence 45, Appl
4	52.8	8.5	7167	US-10-311-455-374	Sequence 374, Appl
5	52	8.4	7128	US-10-311-455-1532	Sequence 1532, Appl
6	51.8	8.3	5523	US-10-311-455-137	Sequence 137, Appl
7	51.8	8.3	73334	US-10-311-455-2097	Sequence 2097, Appl
8	51.2	8.2	12025	US-10-311-455-1272	Sequence 1272, Appl
9	51	8.2	6118	US-10-311-455-1004	Sequence 1004, Appl
10	51	8.2	7921	US-10-311-455-1944	Sequence 1804, Appl
11	50.6	8.1	6126	US-10-311-455-1804	Sequence 1804, Appl
12	50.6	8.1	11047	US-10-311-455-1958	Sequence 1958, Appl
13	50.6	8.1	11047	US-10-240-453-276	Sequence 276, Appl
14	50.6	8.1	11047	US-10-239-676-188	Sequence 188, Appl
15	50.2	8.1	51509	US-09-754-853A-4	Sequence 4, Appl1
16	49.8	8.0	5935	US-10-239-676-133	Sequence 133, Appl

17	49.8	8.0	40862	US-10-311-455-2045	Sequence 2045, Appl
18	48.8	7.8	5542	US-10-311-455-1994	Sequence 1994, Appl
19	48.6	7.8	18855	US-10-311-455-583	Sequence 583, Appl
20	48.2	7.7	12138	US-10-311-455-1602	Sequence 1602, Appl
21	48.2	7.7	12590	US-10-311-455-1893	Sequence 1893, Appl
22	48.2	7.7	18512	US-10-311-455-950	Sequence 950, Appl
23	48	7.7	13427	US-10-311-455-1900	Sequence 1900, Appl
24	47.8	7.7	19972	US-10-311-455-1530	Sequence 1530, Appl
25	47.6	7.7	7110	US-10-311-455-1486	Sequence 1486, Appl
26	47.6	7.7	11092	US-10-311-455-999	Sequence 999, Appl
27	47.4	7.6	6056	US-10-311-455-1454	Sequence 1454, Appl
28	47.4	7.6	6059	US-10-311-455-59	Sequence 59, Appl1
29	47.4	7.6	6385	US-10-239-676-57	Sequence 57, Appl1
30	47.4	7.6	7057	US-10-311-455-1892	Sequence 1822, Appl
31	47.4	7.6	7057	US-10-240-485-148	Sequence 148, Appl
32	47.4	7.6	7057	US-10-311-455-1702	Sequence 1702, Appl
33	47.2	7.6	17721	US-10-311-455-1424	Sequence 1424, Appl
34	47.2	7.6	19787	US-10-311-455-92	Sequence 92, Appl1
35	47	7.5	11416	US-10-311-455-1722	Sequence 1722, Appl
36	46.8	7.5	11805	US-10-311-455-1593	Sequence 269, Appl
37	46.8	7.5	15743	US-10-240-453-269	Sequence 1593, Appl
38	46.6	7.5	6254	US-10-311-455-381	Sequence 381, Appl
39	46.6	7.5	6317	US-10-311-455-2092	Sequence 2092, Appl
40	46.6	7.5	11812	US-10-240-453-306	Sequence 306, Appl
41	46.6	7.5	11812	US-10-239-676-210	Sequence 210, Appl
42	46.6	7.5	11812	US-10-311-455-1810	Sequence 1810, Appl
43	46.6	7.5	6089	US-10-311-455-1515	Sequence 1515, Appl
44	46.4	7.4	6073		
45	46.2	7.4	6073		

## ALIGNMENTS

RESULT 1  
US-10-010-160-1  
Sequence 1, Application US/10010160  
Publication No. US20030103999A1  
GENERAL INFORMATION:  
APPLICANT: Roese, Everett L.  
APPLICANT: Strugnell, Richard A.  
APPLICANT: Good, Robert T,  
APPLICANT: King, Kendall W.  
TITLE OF INVENTION: NOVEL THERAPEUTIC COMPOSITIONS FOR  
TREATING INFECTION BY LAWSONIA SPP.  
FILE REFERENCE: DAVI110.001AUS  
CURRENT APPLICATION NUMBER: US/10/010.160  
CURRENT FILING DATE: 2001-11-09  
PRIOR APPLICATION NUMBER: AU PR1381  
PRIOR FILING DATE: 2000-11-10  
PRIOR APPLICATION NUMBER: US 60/249,596  
PRIOR FILING DATE: 2000-11-17  
NUMBER OF SEQ ID NOS: 68  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 622  
TYPE: DNA  
ORGANISM: Lawsonia intracellularis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1) ... (621)  
US-10-010-160-1

Query Match 100.0%; Score 622; DB 14; Length 622;  
Best Local Similarity 100.0%; Pred. No. 2.3e-123; Indels 0; Gaps 0;  
Matches 622; Conservativeness 0; Mismatches 0;

QY 1 ATGTCTGATGATCCAGTAATAAGAGCAACCCGGAAGCAAGTCTGAGAGTCTGT 60  
DB 1 ATGTCTGATGATCCAGTAATAAGAGCAACCCGGAAGCAAGTCTGAGAGTCTGT 60  
QY 61 TCTGAAGGAGGTCTCCTTAATGAGAGGTTACTTAAGCATTTGACTGACGAGGG 120

Db	61	CTGAAAGGAGATGTCCTCCCTAAATCAGAAAGGTTACTAAAGCATGACTACGACAGAGG	120
OY	121	ATGCGGGGGCTGCGATTTATTCAGCGCGTAATGGGACGTCATTTTGAACAATTTTCTAC	180
Db	121	ATGCGGGGGCTGCTATTTATTCAGCGCGTAATGGGACGTCATTTTGAACAATTTTCTAC	180
OY	181	TATATTTTTCACGAATTCATTTTGGTTGAGTTACAGCACAGTCAGTATATGCTTTATTT	240
Db	181	TATATTTTTCACGATTCATTTTGGTTGAGTTTACAGCACAGTCAGTATATGCTTTATTT	240
OY	241	ATTATGTTGCTCAAGAGATAGCTATTTTATGAGCCAAATTTTACTTTTATTTGCTGT	300
Db	241	ATTATGTTGCTCAAGAGATAGCTATTTTATGAGCCAAATTTTACTTTTATTTGCTGT	300
OY	301	ACGGCATGGAATTCATTCAGTGCACAGTTGGTGCATTATGACTACAAAGGTTTTTAAA	360
Db	301	ACGGCATGGAATTCATTCAGTGCACAGTTGGTGCATTATGAGCTACAAAGGTTTTTAAA	360
OY	361	TTTAAATGAGATTAATTAATTAATTAATAAAGGTTGAAAGAAATGTTGCTTCTACAA	420
Db	361	TTTAAATGAGATTAATTAATTAATTAATAAAGGTTGAAAGAAATGTTGCTTCTACAA	420
OY	421	ACACTTGTTCGACTTTTACGTAGTTAGTTCAGTAATGTATATGATTTGTTCCATAT	480
Db	421	ACACTTGTTCGACTTTTACGTAGTTAGTTCAGTAATGTATATGATTTGTTCCATAT	480
OY	481	ATGATTAATAAAGGAGAGTTTTCAAACTTTTACCATATATATATGCAAGTCTTCAGGT	540
Db	481	ATGATTAATAAAGGAGAGTTTTCAAACTTTTACCATATATATATGCAAGTCTTCAGGT	540
OY	541	GTGGCAGATTATGCTTAATAACAGGAATAGACTGTTTATATATACCTAATTCCTATG	600
Db	541	GTGGCAGATTATATGCTTAATAATACAGGAATAGACTGTTTATATATACCTAATTCCTATG	600
OY	601	ACAATTAATTCAGTGCAGATC 622	
Db	601	ACAATTAATTCAGTGCAGATC 622	

```

RESULT 2
US-10-311-455-118
Sequence 118, Application US/10311455
Publication No. US20030143606a1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining the Cytosine Methylation Level of a Specific Cytosine
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311,455
PRIORITY FILING DATE: 2002-12-16
PRIORITY APPLICATION NUMBER: PCT/EP01/07537
PRIORITY FILING DATE: 2001-07-02
PRIORITY APPLICATION NUMBER: DE 10032529.7
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: DE 10043826.1
PRIORITY FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 118
LENGTH: 9052
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-118

```

```

Query Match      8.8%; Score 54.8; DB 12; length 9052;
Best Local Similarity 47.2%; Pred. No. 0.083;
Matches 167; Conservative 0; Mismatches 187; Indels 0; Gaps 0

138 TTATTACGGCGCTAATGGGACGTCATTTTGGAAACAATTTTCTACTATTTTACAGAAATC 197
|| || || || || || || || || || || || || || || || || || || || || ||

```

Db 799 TTTTATAGAGAAAAAGATTGATTATTAATTTAGTTTGTCTTTTATTTATTTGTTAT 858

Qy 198 ATTTCGGTTAGGTTACAGCAGCAGTACGATATGCTTATTTATTTATTTATGTCACAGA 257

Db 859 TTATTTTTTTATTTATTTTATTTATTTGTTATTTATTTATTTATTTATTTATTTATTTA 918

Qy 258 GATAGCTATTTTATTTGATGATCCAAATATTACTTTTATTTGCTGTACGCGATGGATTCCAT 317

Db 919 TTTGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 978

Qy 318 ACGGTACAAGTGGTGCAATACGACACAAAGGTTTTAAATTTAAATGAGAAATTT 377

Db 979 ATTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1038

Qy 378 TTAATTAATAAAGGCGTTGAAAGGAATGTTTGCCTTCACAAACACTGTTCGACTTTT 437

Db 1039 TGTGAAATAATACGATTTTATGATGAGGTATTTGTGAAGTATTTTTTTTTTGAGTAAG 1098

Qy 438 ACGTAGTTTACTTCAAGTAATGTTATTAAGGATTTGCCAATAATGATTTATAAA 491

Db 1099 AGGTATAGGATTTTTTGCATTTTTTTTGTCTTTATTTTAAATGTTTGGTGATGAGA 1152

```

RESULT 3
US-09-895-913A-45/C
; Sequence 45 Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
; APPLICANT: Kleantous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020160456A1 Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/043002
; CURRENT APPLICATION NUMBER: US/09/895,913A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 08/881,227
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 945
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (61) ... (636)
; US-09-895-913A-45

```

Query Match	8.5%	Score 53	DB 10	Length 945
Best Local Similarity	46.9%	Pred. No. 0.084		
Matches 201	Conservative 0	Mismatches 225	Indels 3	Gaps 1

  

QY	19	AAAACAGAGAAAGCAACCCCGAAACGACGTCAGAGACGCTGTTTGAAGGAGTGTCCT	78
		AAAAA	
Db	845	AAAAACCCAMCTCCCTAGCGCGAAAAAATAATCCAAAAAACCCAGAGAAAGCAATGTGCT	786
QY	79	AAATCAGAAAGGTTACTAAGAATTCATCTGCACAGAGAGTGGGGCTGCTAT	138
		785	
Db	785	AAGAGCATGGAAGTGTGTGGGGTTTTTGGGTTATTGGCCGGCTAATTAGATTTTTGTT	726
QY	139	TATTCAGCGCTAAATGGACGTCATTTTGAACAATTTTCTACTATATTTTACAGATCA	198
		725	
Db	725	TTTTTATATGATGGGTGATGAGCTTTAGCGAAATGATGCGCATGTGTGAAGA---T	669
QY	199	TTTTCGTTTGAAGTTTACAGACAGTACGATATATCTTATTTATTTATTTGTTCTCAAG	258
		668	
Db	668	TTTTTCCTAATTTCACTAAGAAAGGCTTCAGAGCTGTTTATCCAACTGCTGTTAAGAC	609
QY	259	ATAGCTATTTTATGATGACCAATATTACTTTTATATGCTGTACGGCATGGATTTTCATTA	318

Db 608 ACTTTTATGCTTTTACCATTATTTATCATTTTAGGGGGGGCCCTTTTATCTAAT 549  
QY 319 CGGTACAGATTGGTGCATTAAGCACTAAGCTTTTAAATTTAAATGAGTAATTT 378  
Db 548 GCTTGCATTTTGGCTGGCTCTTTGGCCCTTAAGTCATGAGCTTAATTTTCTAAATC 489  
QY 379 AATAATATAAGGCTGAAGAAAGAAATGTTTCTTCAACAACAACCTTGTGACTTTA 438  
Db 488 AACCTTATCAATGGCTCAAAACCTTTTCTTAAAAAAGCTCCTTATGAGGAGTTTG 429  
QY 439 CGTAGTTTA 447  
Db 428 ATCACCCTTA 420

RESULT 4  
US-10-311-455-374  
; Sequence 374, Application US/10311455  
; Publication No. US20030143606A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect  
; FILE REFERENCE: 5013.1014  
; CURRENT APPLICATION NUMBER: US/10/311.455  
; PRIOR FILING DATE: 2002-12-16  
; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 2424  
; SEQ ID NO 374  
; LENGTH: 7167  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-374

Query Match 8.5%; Score 52.8; DB 12; Length 7167;  
Best Local Similarity 46.3%; Pred. No. 0.2;  
Matches 174; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

QY 234 TTATTTATTTTANGTCTCAAGATAGCTATTTTATGATGCAATATTTACTTTTAT 293  
Db 709 TATATTTGTTAGTATTAGAGATTTTAAGTATAGATTTATATTTATTTTATTTAG 768  
QY 294 TGCCTTACGCGATTTTCAATACGTACAAAGTTGGTCAATTAATGACTACAAAGT 353  
Db 769 AGATTTAGTCTTATGATTTTATATTTTATATTTTAAATGTTAGGCTTTTGAAT 828  
QY 354 TTTTAAATTTTAAATGAGTAATTTAATATAAAGGTTGAAGGAATGTTGCTC 413  
Db 829 ATTATTTGTTAGATTAATTAATTTTAAATTTAAGGTTGAAGTTTAAATTT 888  
QY 414 TCAACAACAACCTTGTGACTTTTACGATTTTACGTTCACTTAATTTTATAGGATTTG 473  
Db 889 TGCTTTATATATGTTTAAATGTTTAGGGCTTTGATGCTATTTGTTAAAGTTT 948  
QY 474 TCCATATATGTTTAAAGAGAGTTTCAACCTTTTACATTAATTAATGCAAGTCC 533  
Db 949 AATAATTTGGGCTTTTAAAGCAATTTTAAATTTTAAATTTTAAAGTTTATGATTA 1008  
QY 534 TTCAGTGTGGCATTAATGCTTAATACAGAAATGACTTGTTTTATTAAGCTAAT 593  
Db 1009 TTTTGAAGTTTAAAGAAATTTATATTTTATGATTAATTTATGTTTTCGATTAATTTATG 1068  
QY 594 TCCATGACAAATTAAT 609

Db 1069 TTTTAAAGAAATTTT 1084

RESULT 5  
US-10-311-455-1532  
; Sequence 1532, Application US/10311455  
; Publication No. US20030143606A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect  
; FILE REFERENCE: 5013.1014  
; CURRENT APPLICATION NUMBER: US/10/311.455  
; PRIOR FILING DATE: 2002-12-16  
; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 2424  
; SEQ ID NO 1532  
; LENGTH: 7128  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-1532

Query Match 8.4%; Score 52; DB 12; Length 7128;  
Best Local Similarity 46.5%; Pred. No. 0.3;  
Matches 240; Conservative 0; Mismatches 270; Indels 6; Gaps 2;

QY 78 TAAATCAGAAAGAGTTACTTAAGCATTCAGTACGACGAGGATGCGGCTTGCTAT 137  
Db 4809 TAAATTTATGATTAATTAATTTATTTATTTTAAATGATTAATGATTAAT 4868  
QY 138 TTATTCAGCGTTAATGGAGCTCATTTGAAACAATTTTCTACTATTTTACGAATC 197  
Db 4869 TTAGTTAGTATGATGATTAATTTTATTTTATTTTAAATTTTATTTTAAAT 4928  
QY 198 ATTTGCGTTGAGGTTACAGACAGACAGATATGCTTTATTTATGTTGCTCAAGA 257  
Db 4929 TTAAATATGATTAATTAATTAAGTTTATTAATGATTAATGATTTTATGTTG 4988  
QY 258 GATAGCTATTTTATGATGCAATATTAATTTTATTTGCTTTACGCGATTTGAT 317  
Db 4989 TAT-----TATGATGCAAAATTTTGAAGTTTTTTATGAGTTAATTTGATTT 5043  
QY 318 ACGTGTACAGTTGCTGATTAATGACTACAAAGTTTAAATTTAAATGAGTAAT 377  
Db 5044 TAGTTTATTTTATGATTTTATTTTATTTTAAATTTTATTTTAAATTTTATG 5102  
QY 378 TAAATTAATAAAGGTTGAAGGAATGTTGCTTCAACAACAACCTTGTGACTTT 437  
Db 5103 TTGATATGTTTATGATTTATTTAGGATTAATTAATTAATTAATTAATTTATTTAT 5162  
QY 438 ACGTAGTTTATGCTCAAGTAATTTGTTATAGTATTTGCTCATTAATTTAAAGAGA 497  
Db 5163 GGGGTTATTTTATGATTAATTTATTTTATTTTAAATTTTAAATTTTAAAGAGAAT 5222  
QY 498 GTTTCAACTTTTACATTAATTAATGCAAGTCTTCAAGGTGCGCAATTAATGCT 557  
Db 5223 GATTTTATGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAT 5282  
QY 558 TAAATGCAATGATGCTTTTATTAATGCTAAT 593  
Db 5283 TAAATTTAAATTAATTAATTAATTAATTTATTTATTT 5318

```
RESULT 6
US-10-311-455-137
; Sequence 137, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 137
; LENGTH: 5523
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-137

Query Match
Best Local Similarity 45.6%; Score 51.8; DB 12; Length 5523;
Matches 182; Conservative 0; Mismatches 217; Indels 0; Gaps 0;

QY 185 TTTTACACAAATCATCTTCGTTGAGGTACAGACAGACAGTATATGCTTATTTATTT 244
DB 726 TTTTATAAATTTTGTGTTGAGGTGTTGATTTAGTATATGTTTAAATAGTT 785
QY 245 ATGTTGCTCAAGATAGCTATTTTATGATGCATATTTACTTTTATGCTGTTACG 304
DB 786 ATGAGTTTGAAGTTTAAATATTTATGCTTTGTTTGTATTTTAAATTTTATAG 845
QY 305 CATGATTTCAATTCAGTACAAAGTTGTCATTTATGACATACAAAGTTTAAATTTA 364
DB 846 AATTAAAGATGATGAGTGAATGGGTTAAGAGATCGTTTATTTGTTTATTTATAT 905
QY 365 AATGAGTAAATTTATATATATATATATATATATATATATATATATATATATAT 424
DB 906 ATAGATGAGGATGTTATTTATAGTCGTAAGAAATTTGTTTATGTTTATAGAT 965
QY 425 TTGTTGCACTTTAGCTAGTTTCAAGTATGTTATAGGATTTGTTTCAATATATGA 484
DB 966 TTTTATTTTATTTTATAGTTTATGTTTATGTTATTTTAAATGAAATGTAATATA 1025
QY 485 TTATAAAGAGAGATTTCAAACTTTTACATATATATATATATATATATATATAT 544
DB 1026 GATGATATATTTTAAAGATATATATATATATATATATATATATATATATAT 1085
QY 545 CAGATATATGCTTATATATATATATATATATATATATATATATATATATATAT 583
DB 1086 TATATATATATATATATATATATATATATATATATATATATATATATATAT 1124

RESULT 7
US-10-311-455-2097
; Sequence 2097, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
```

```
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2097
; LENGTH: 73334
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2097

Query Match
Best Local Similarity 49.8%; Score 51.8; DB 12; Length 73334;
Matches 131; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 264 TATTTATTTATGATCCAAATTTACTTTTATTTGCTGTACGGCATGATTTCAATGCT 323
DB 5766 TAAATATTTAAATGGAATAGAAATTTTATTTGATTTTAAAAAAGGTTA 5725
QY 324 ACAATGTCGATATGAGCTACAAAGTTTAAATTTAAATGAGTAAATTAATAT 383
DB 57726 TAAATATGAGGTAATATGATTTATTTTGTGAAGTGAAGTAAGAAATTTATAT 57785
QY 384 AATAAAGGTTGAAGAAATGTTTCTTCAACAAACATTTGTCGATTTTACGTAG 443
DB 57786 TTTATATTAATTTAAATATGTTATTTATTTATTAAGAAAGAAATTTTATAT 57845
QY 444 TTTAGTTCAAGTATTTGTTATAGGATTTGTCATATATGATTTAAAGAGAGTTTC 503
DB 57846 TTAATGGAATTTGATATATATTTTATTTATTTATTTATTTATTTATTTAGGGGTTT 57905
QY 504 AAATTTTACATTAATATATAG 526
DB 57906 AATATTTTGTATATAGTTATAG 57928

RESULT 8
US-10-311-455-1272
; Sequence 1272, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1272
; LENGTH: 12025
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1272

Query Match
Best Local Similarity 46.5%; Score 51.2; DB 12; Length 12025;
Matches 201; Conservative 0; Mismatches 228; Indels 3; Gaps 1;
```



QY 163 TTGAAACATTTCTACTATATTTTACAGAAATTCGTTGAGTTACAGCAG 222  
D 203 TTTTAGAGATTTAGTTTGTATTTTAAATTTATTCAGATGATGTTAGTTGAATTA 262  
QY 223 TCATATATGCTTATTTATTTATTTGCTCAAGATAGCTATTTTATTCAGTCCATA 282  
D 263 TAGAATATTTGGGTTTTTTTTTTTAAATGTTATTTGATTTTTTTTAAATATGATA 322  
QY 283 TTACTTTTATTTGCTGTACGAGATTCATTTACGTACAAAGTTGTCATTTAG 342  
D 323 TGATTTTATAGAGAGATTTTATAGAAATTCGTTGATTAATTTATATGATTTTGTGA 382  
QY 343 ACTACAAAGTTTAAATTTAAATGAGTAATTTATATATATTAAGGGTTGAAAGA 402  
D 383 ATTAACGATATTTAAATATGATTAATGTTAAAGTATTTAGTGTAGAGG 442  
QY 403 ATGTTGCTCTCAACAAACCTTG--TTGACCTTTAGCTAGTTAGTTCAAGTAAT 459  
D 443 AAGTTTTTAAATTTGATTTTGAATTTTGAATTTTAAAGAAATTTTTTGTGTAATTTT 502  
QY 460 GTTATAGGATTTGTCATATATATTAATAAGAGAGTTTCAACTTTTACATTA 519  
D 503 TATTAATGATTTATTTAAATTTTATGATTAAGATTTTATTTAGCTTTGATATATGTA 562  
QY 520 TATTAATGCAAGTCTTCAGGTGTGCAAGATTTATGCTTAATACAGATGATCTTGT 579  
D 563 TTTTAAATTTAGTTGTGTTGCAATTAATAATGTTTATATATGAAATTTATGATAGGT 622  
QY 580 TTATATACGCTA 591  
D 623 TTAAAGTTGTA 634

## RESULT 9

US-10-311-455-1004  
Sequence 1004, Application US/10311455  
Publication No. US20030143606A1  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENBROCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detecting  
FILE REFERENCE: 5013.1014  
CURRENT APPLICATION NUMBER: US/10/311.455  
CURRENT FILING DATE: 2002-12-16  
PRIOR APPLICATION NUMBER: PCT/EP01/07537  
PRIOR FILING DATE: 2001-07-02  
PRIOR APPLICATION NUMBER: DE 10032529.7  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: DE 10043826.1  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 2424  
SEQ ID NO 1004  
LENGTH: 6118  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-1004

Query Match 8.2%; Score 51; DB 12; Length 6118;  
Best Local Similarity 50.5%; Pred. No. 0.46; Indels 3; Gaps 1;

Matches 151; Conservative 0; Mismatches 145; Indels 3; Gaps 1;

QY 227 TATATGCTTATTTATTTATTTATGCTCAAGATAGCTATTTTATGATGCAATATTTAC 286  
D 4654 TAAATAGCTATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTT 4713  
QY 287 TTTTATTTGCTTACGCAATGATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 346  
D 4714 TTTTATTTGCTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 4770

QY 347 CAAGGTTTTTAAATTTAAATGAGTAATTTATATATTAATAAGGGTTGAAAGAAATGT 406  
D 4771 GAGATTTTAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATG 4830  
QY 407 TTGCTTCTCAACAAACCTGTTGCACTTTTACGATTTAGTTAGTTCAAGTAATTTGTATAG 466  
D 4831 AAGTTCATTTATTTGGGATTTAGTTTATTTATTTATTTATTTATTTATTTATTTATTT 4890  
QY 467 GTATTTGCTCATATATGATTTATTAATAAGAGAGTTTCAACTTTTACATTTATTTAT 525  
D 4891 TTATAGTTTATTTTATTTTGAATAATGAAATATGATTAATATATATATATATTTATAGTGT 4949

## RESULT 10

US-10-311-455-1944  
Sequence 1944, Application US/10311455  
Publication No. US20030143606A1  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENBROCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detecting  
FILE REFERENCE: 5013.1014  
CURRENT APPLICATION NUMBER: US/10/311.455  
CURRENT FILING DATE: 2002-12-16  
PRIOR APPLICATION NUMBER: PCT/EP01/07537  
PRIOR FILING DATE: 2001-07-02  
PRIOR APPLICATION NUMBER: DE 10032529.7  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: DE 10043826.1  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 2424  
SEQ ID NO 1944  
LENGTH: 7921  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-1944

Query Match 8.2%; Score 51; DB 12; Length 7921;  
Best Local Similarity 46.5%; Pred. No. 0.51;

Matches 165; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

QY 159 TCATTTGAAACAATTTCTACTATATTTTACAGAAATTCGTTGAGTTACAGC 218  
D 2153 TGAATTTTACAGCAAGTATTTAGTTGTTAGTAAATTTTGTATTTATTTTATTTTACGGT 2212  
QY 219 ACAGTCAATATGCTTATTTATTTATTTGCTCAAGATGATGCTATTTATTTATGATGCC 278  
D 2213 TGTGGGGGATTTAGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 2272  
QY 279 AATATTAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 338  
D 2273 AGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 2332  
QY 339 ATGACTACAAAGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 398  
D 2333 TGGGAATTAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 2392  
QY 399 AGGAATTTTGTCTTCAACAAACCTGTTGCACTTTTACGATTTTATTTATTTATTTATTT 458  
D 2393 AAGATGCTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 2452  
QY 459 TGTATATGATTTTGTCTCATATATGATTTATTAATAAGAGAGTTTCAACTTTTATTA 513  
D 2453 TATATATGCTTCAATTTTATTTTATTTTGTGGATGTAATTTGTTATTTATTTATTTATTT 2507

## RESULT 11

US-10-311-455-1804  
Sequence 1804, Application US/10311455

Query Match	8.1%	Score 50.6	DB 12	Length 6126
Best Local Similarity	46.7%	Pred. No. 0.56		
Matches 161	Conservative	0	Mismatches 184	Indels 0
				Gaps 0

RESULT 12  
US-10-311-455-1958

```

: NUMBER OF SEQ ID NOS: 2424
: SEQ ID NO 1958
: LENGTH: 11047
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1958

```

Query Match	8.1%;	Score 50.6;	DB 12;	Length 11047;
Best Local Similarity	46.9%;	Pred. No. 0.71;		
Matches 158;	Conservative 0;	Mismatches 179;	Indels 0;	Gaps 0;

RESULT 13  
US-10-240-453-276

Query Match 8.1%; Score 50.6; DB 12; Length 11047;

Best Local Similarity 46.9%; Pred. No. 0.71;  
Matches 158; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

QY 164 TTGAACAATTTTCTACATATTTTACAGATCATTTGCGTTAGGTTACAGACAGT 223  
Db 10382 TTGAATGAAATTAAGATTAATTTTGAATATATATTTGTTTGTAGTTAGAGGA 10441  
QY 224 CAGTATGCTTATTTATTTATTTATGTTGCTCAAGATATGATTTATTTATGATTCATAT 283  
Db 10442 GCGATTTATTTATTTATTTATTTATTTGTTGAATTTTATTTTATTTTGTGAAGTTGG 10501  
QY 284 TACTTTTATTTGCTGTTACGCGATGATTTCAATTCATGTTACCAAGTTGTCATTTAGSA 343  
Db 10502 TTTTATTTTAAAGTAAATGATTAAGATTAAGATTAAGATTAAGATTAAGATTTT 10561  
QY 344 CTACAAAGCTTTTAAATTTTAAATGAGTAAATTTATATTAATTAAGGTTGAAGGAA 403  
Db 10562 TTGATTTATTTTATTTTATTTTAAATTTATTTTTCGCTTAAAGTTGTAATTTATTA 10621  
QY 404 TGTTCCTTCTCAACAAACCTTGTTCGACTTTTACGTAGTTAGTTGCAAGTAAATTTGTTA 463  
Db 10622 TTTTATTTTATTTATTTATTTATTTTGTGTTAGTTAGTTAGTTAGTTAGTTAGTT 10681  
QY 464 TAGTATTTGTTCCATATATATGATTAATTAAGAGAGGTT 500  
Db 10682 TTGATGTAATTTTATTTAGATTTAGTTATTTGGGAATTT 10718

RESULT 14  
US-10-239-676-188  
; Sequence 188, Application US/10239676  
; Publication No. US20030082609A1

GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENBROCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
FILE REFERENCE: 5013.1003  
CURRENT APPLICATION NUMBER: US/10/239,676  
CURRENT FILING DATE: 2002-09-24  
PRIORITY APPLICATION NUMBER: PCT/EP01/03968  
DE 10019058.8  
DE 10019173.8  
DE 10032529.7  
DE 10043826.1  
PRIORITY FILING DATE: 2001-04-06  
2000-04-06  
2000-04-07  
2000-06-30  
2000-09-01  
NUMBER OF SEQ ID NOS: 228  
SEQ ID NO 188  
LENGTH: 11047  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-239-676-188

Query Match 8.1%; Score 50.6; DB 14; Length 11047;  
Best Local Similarity 46.9%; Pred. No. 0.71;  
Matches 158; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

QY 164 TTGAACAATTTTCTACATATTTTACAGATCATTTGCGTTAGGTTACAGACAGT 223  
Db 10382 TTGAATGAAATTAAGATTAATTTTGAATATATATTTGTTTGTAGTTAGAGGA 10441  
QY 224 CAGTATGCTTATTTATTTATTTATGTTGCTCAAGATATGATTTATTTATGATTCATAT 283  
Db 10442 GCGATTTATTTATTTATTTATTTATTTGTTGAATTTTATTTTGTGAAGTTGG 10501  
QY 284 TACTTTTATTTGCTGTTACGCGATGATTTCAATTCATGTTACCAAGTTGTCATTTAGSA 343  
Db 10502 TTTTATTTTAAAGTAAATGATTAAGATTAAGATTAAGATTAAGATTAAGATTTT 10561  
QY 344 CTACAAAGCTTTTAAATTTTAAATGAGTAAATTTATATTAATTAAGGTTGAAGGAA 403  
Db 10562 TTGATTTATTTTATTTTATTTTAAATTTATTTTTCGCTTAAAGTTGTAATTTATTA 10621  
QY 404 TGTTCCTTCTCAACAAACCTTGTTCGACTTTTACGTAGTTAGTTGCAAGTAAATTTGTTA 463  
Db 10622 TTTTATTTTATTTATTTATTTATTTTGTGTTAGTTAGTTAGTTAGTTAGTTAGTT 10681  
QY 464 TAGTATTTGTTCCATATATATGATTAATTAAGAGAGGTT 500  
Db 10682 TTGATGTAATTTTATTTAGATTTAGTTATTTGGGAATTT 10718

Db 10502 TTTTATTTTAAAGTAAATGATTAAGATTAAGATTAAGATTAAGATTAAGATTTT 10561  
QY 344 CTACAAAGCTTTTAAATTTTAAATGAGTAAATTTATATTAATTAAGGTTGAAGGAA 403  
Db 10562 TTGATTTATTTTATTTTATTTTAAATTTATTTTTCGCTTAAAGTTGTAATTTATTA 10621  
QY 404 TGTTCCTTCTCAACAAACCTTGTTCGACTTTTACGTAGTTAGTTGCAAGTAAATTTGTTA 463  
Db 10622 TTTTATTTTATTTATTTATTTATTTTGTGTTAGTTAGTTAGTTAGTTAGTTAGTT 10681  
QY 464 TAGTATTTGTTCCATATATATGATTAATTAAGAGAGGTT 500  
Db 10682 TTGATGTAATTTTATTTAGATTTAGTTATTTGGGAATTT 10718

RESULT 15  
US-09-754-853A-4/c  
; Sequence 4, Application US/09754853A  
; Publication No. US20030005491A1

GENERAL INFORMATION:  
APPLICANT: Hauge, Brian M.  
APPLICANT: Parneil, Laurence D.  
APPLICANT: Wang, Ming Li  
APPLICANT: Parsons, Jeremy D.  
TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
FILE REFERENCE: 38-10(115810)B  
CURRENT APPLICATION NUMBER: US/09/754,853A  
CURRENT FILING DATE: 2001-01-05  
PRIORITY APPLICATION NUMBER: US 60/174,880  
PRIORITY FILING DATE: 2000-01-07  
NUMBER OF SEQ ID NOS: 1119  
SEQ ID NO 4  
LENGTH: 513509  
TYPE: DNA  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (111805)..(113968), (114684)..(115204)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(513509)  
OTHER INFORMATION: unsure at all n locations  
OTHER INFORMATION: Clone ID: 318013\_region\_A3  
US-09-754-853A-4

Query Match 8.1%; Score 50.2; DB 11; Length 513509;  
Best Local Similarity 47.4%; Pred. No. 3.8;  
Matches 213; Conservative 0; Mismatches 233; Indels 3; Gaps 2;

QY 147 CGTATGAGGAGCTATTTTGAACAATTTCTACTATATTTTAA--CAGATTCATTTGG 204  
Db 182633 CGTATGAGGAGGAAATGATTTTATACAAATTTGTATGACTTTATTTTAAAGTAAATTTGATA 182574  
QY 205 TTGAGGTTACAGACAGTCAGTATATGCTTATTTATTTATTTATTTGCTCAAGATAGCT 264  
Db 182573 ATTTTATGCAATTAAGATTAATTTAAACTATTTTCTTACTAATCAATTAACAAAC 182514  
QY 265 ATTTTATGATGCCAATATTAATTTTATTTGCTGTTACGCGATGATTTCAATTCATTCGTTA 324  
Db 182513 TATTTATATATGATTAATGATTAATTTTAAACCTATTTTGTACCGGTTGTTACGAATA 182454  
QY 325 CAAGTTGTCATTAATGACCTACAAAGTTTAAATTTTAAATTTAAATGAGATTAATTTAATATA 384  
Db 182453 AAAAATGACATTTTATTTTACAGTTATTTATTTATTTAAAGATTAATTAAGATTAATA 182394  
QY 385 ATTAAGGTTGAAGAAATGTTGCTTCAACAAACCTTTGCACTTTTACGTTAGT 444  
Db 182393 ATTAAGAAATTAATTAATTAATTAATTAAGAAAGAAATGTTAGAAATATATTTTAT 182334  
QY 445 TTAGTTCAAGTATTTTATTAAGTATTTGTTCAATATATGATTAATTAAGAGAGGTTTCA 504  
Db 182333 TTTTATTTATTTGATTAATTTCAATTAATTAATTAATTAATTTTATTTTATTTTATTTTACTT 182275

QY 505 AACTTTTACCATTAATTAATGCAAGTCTTCAAGTGTGCGACATTAATATGCTTAATAC 564  
DB 182274 AATTATAGCGCATTTTATTTTACCCTCATTCACATTAATTAACATTAATACCA 182215  
QY 565 GGATAGTACTGTTTATTAATACGCTAAT 593  
DB 182214 ATTTATGCAAAATATTTTCTCTATCTAAT 182186

## RESULT 16

US-10-239-676-133  
; Sequence 133, Application US/10239676  
; Publication No. US20030082609A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
; FILE REFERENCE: 5013.1003  
; CURRENT APPLICATION NUMBER: US/10/239,676  
; PRIOR FILING DATE: 2002-09-24  
; PRIOR APPLICATION NUMBER: PCT/EP01/03968  
; DE 10019058.8  
; DE 10019173.8  
; DE 1003529.7  
; DE 10043826.1  
; PRIOR FILING DATE: 2001-04-06  
; 2000-04-06  
; 2000-04-07  
; 2000-06-30  
; 2000-09-01  
; NUMBER OF SEQ ID NOS: 228  
; SEQ ID NO 133  
; LENGTH: 5935  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-239-676-133

Query Match 8.0%; Score 49.8; DB 14; Length 5935;  
Best Local Similarity 47.3%; Pred. No. 0.82;  
Matches 183; Conservative 0; Mismatches 202; Indels 2; Gaps 1;

QY 227 TATATGCTTATTTATTTATGTTGCTCAAGATACCTTTTATTTGATGCCAATATTC 286  
DB 2608 TTTAAGATTTTATTTAGTGTGTTTGAGAGTGTGTTTATTTATTTATTAATTAAT 2667  
QY 287 TTTTATGCTGTAGCGATGATTTCAATTAAGTGTGATGATTAATGAGCTA 346  
DB 2668 TTTAATAGTAAATTTGTTATGTTTT--TTAATTAATGAATTTGTGAGAAATTAAT 2725  
QY 347 CAAAGTTTAAATTAATTAAGAGTAAATTAATTAATTAATTAAGGTTGAAAGAAATGT 406  
DB 2726 ATGAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2785  
QY 407 TTGCTTCGAACAAACACTGTTGCACTTTAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 466  
DB 2786 TAGTATTTAATTAATTTGTTAATTTTATTAATTAATTAATTAATTAATTAATTTT 2845  
QY 467 GTATGTTCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 526  
DB 2846 ATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTGTG 2905  
QY 527 CAAATCTTCAGGTGTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 586  
DB 2906 GTTGAATTTGAATTTATTTTGTGTTTATTAATTAATTAATTAATTAATTAATTAAT 2965  
QY 587 CGCTAATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 613  
DB 2966 TTGATTAATTAATTAATTAATTTTGTG 2992

## RESULT 17

US-10-311-455-2045  
; Sequence 2045, Application US/10311455  
; Publication No. US20030143606A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ  
; FILE REFERENCE: 5013.1014  
; CURRENT APPLICATION NUMBER: US/10/311,455  
; PRIOR FILING DATE: 2002-12-16  
; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 2424  
; SEQ ID NO 2045  
; LENGTH: 40862  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-2045

Query Match 8.0%; Score 49.8; DB 12; Length 40862;  
Best Local Similarity 45.2%; Pred. No. 1.7;  
Matches 183; Conservative 0; Mismatches 222; Indels 0; Gaps 0;

QY 108 TACTGACAGAGATGCTGGGCTTCTATTTATTCAGCGGTAATGAGACGTCATTTGA 167  
DB 16338 TATTGGGAATGATTAATTAATTTGTTTATTAATTTAGTGTGAATTAATGTTAT 16397  
QY 168 AACATTTCTACTATTTTATTTTACAGATCAATTCGTTGAGTTACAGACACTGCT 227  
DB 16398 CGATTAATTTATTTGTTTGTATGATTAATTAATTAATTAATTAATTAATTAATTA 16457  
QY 228 ATATGCTTATTTATTTATTTGTTGCTCAAGATAGCTTTTATTTATGATGCCAATTTACT 287  
DB 16458 AATTAATTTTATTTATTTATTTGTTGTAATTAATTAATTAATTTGTTGTTTAT 16517  
QY 288 TTTTATGCTGTAGCGATGATTTCAATTAAGTGTGATTAATTAATTAATTAATTAAT 347  
DB 16518 TGGGTTTTTTTTTTTAAAGATGAATTAATTAATTAATTAATTAATTAATTAATTA 16577  
QY 348 AAAGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 407  
DB 16578 AATTTAGGTTTATTTAGTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 16637  
QY 408 TGTCTTCGAACAAACACTGTTGCACTTTAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 467  
DB 16638 TATTAATTTTAAAGAAATTTTATTAATTAATTAATTAATTAATTAATTTTTCGT 16697  
QY 468 TATGTTCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 512  
DB 16698 TAGTTTGTGAATTAATTAATTTTGTGTTTATTAATTAATTAATTAATTAATTTT 16742

## RESULT 18

US-10-311-455-1994  
; Sequence 1994, Application US/10311455  
; Publication No. US20030143606A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ  
; FILE REFERENCE: 5013.1014

```

1 CURRENT APPLICATION NUMBER: US/10/311,455
2 CURRENT FILING DATE: 2002-12-16
3 PRIOR APPLICATION NUMBER: PCT/EP01/07537
4 PRIOR FILING DATE: 2001-07-02
5 PRIOR APPLICATION NUMBER: DE 10032529.7
6 PRIOR FILING DATE: 2000-06-30
7 PRIOR APPLICATION NUMBER: DE 10043826.1
8 PRIOR FILING DATE: 2000-09-01
9 NUMBER OF SEQ ID NOS: 2424
10 SEQ ID NO 1994
11 LENGTH: 5542
12 TYPE: DNA
13 ORGANISM: Artificial Sequence
14 FEATURE:
15 OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
16 US-10-311-455-1994

```

Query Match	7.8%;	Score 48.8;	DB 12;	Length 5542;
Best Local Similarity	54.8%;	Pred. No. 1.3;		
Matches 160;	Conservative 0;	Mismatches 127;	Indels 5;	Gaps 3

0y 236 GMAATATGCTTATTTATTTATTTGTCCTCAAGAGATAGCAATTTTATTTGATGCCAATATTA 285  
 2468 GTATATTTGCTATTTGTTATTTTTCGAGTAATGTTGTTTACATAT--TTGTGTT 2522

QY 286 CTTTATTATGCTGTACCGCATGGATTTCATTAGCAGAAAGTGCGTAGTAATGACT 345  
Db 2525 ATTTTTAAATTTGTTGTTGTTTTTAATGATGTGAAGAATTATAATATTTAGA-T 2588

OY 346 ACAAAAGTCTTTTAAATTTTAAATGAGTAAATTTATATATAAAAGGTCGAAGCAATG 405  
 Db 2584 ATAGAAATTTTAAATACGATTTTGTAAATATTTTTTGTAAATTTGTGTAACTTAT- 2644

4b TTGCTCTCACAACACACTGTGTGACATTTCACAGTTAGTTCACAGTAATGTATA 465  
 4c TTGCTCTCACAACACACTGTGTGACATTTCACAGTTAGTTCACAGTAATGTATA 465  
 Db 2643 TTGCTTTGTTAAAAATATTTTTTAAGATATATATAATGATATATAAAGATTATTTTAA 2702

2703 GTGATTTTATATATATGATTTTATATGAAAAATTAAAAATTAT 2754  
 458 GGAATGATTCATATATGATTAATAAGGAGATTTTCAACATTTACAT 517  
 2704

RESULT 19  
US-10-311-455-583  
; Sequence 583, Application US/10311455

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

1  TITLE OF INVENTION:  Diagnosis of Diseases Associated with the Immune System
2  TITLE OF INVENTION:  Cytosine methylation
3  FILE REFERENCE:  5013.1014

```

CURRENT FILING DATE: 2002-12-16  
PRIOR APPLICATION NUMBER: PCT/EP01/07537  
PRIOR FILING DATE: 2001-07-02

PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: DE 10043826.1  
PRIOR FILING DATE: 2000-09-01

```

: SEQ ID NO 583
:
: LENGTH: 18855
:
: TYPE: DNA
:
: ORIGIN: 1..18855

```

```
; FEATURE: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION:
US-10-311-455-583
```

Query Match	7.8%;	Score 48.6;	DB 12;	Length 18855;
Best Local Similarity	46.9%;	Pred. No 2.3;		
Matches 187; Conservative	0;	Mismatches 209;	Indels 3;	Gaps 1

Oy	125	TGGGCGTTCGTATTATTTACAGCGATGGAGCGATTCTGAACAATTTCTACTATA	184
Dd	10618	TAGGATTAATTTTTTATTATTATTTTTTAGGGGTAATTTTTTTTTTAGTAGATAATTTA	10677
Oy	185	TTTTTACAGAATCATYTCGGTTTGAGTTACAGCACAGTCAGTANAGCTTATTAATTT	244
Dd	10678	GTTTTTAAAAAGTTTTATGTTTTTGTTTTTATAAAATATGGGGTTTTTTTTTATTTGTT	10733
Oy	245	AAGTTGCTCAAGAGATAGCTATTTTANTGATGCCAATATTACTTTTATTTGCTGTACGG	304
Dd	10738	ATTATTAATGATTAATAATTAATTTTTTTTTTATTTAATTAATGATGTTAGTTGGTAAT	10797
Oy	305	CATGCAATTTCAATACGTCACAAGTGG--GTGCATTATGCACTACAAAGGTTTTAAAT	361
Dd	10798	TTTGCGTAATGTTTTTATTATTTAATTTAAATNTAGTGTTGAAAATTAATAATTTAAAG	10855
Oy	362	TTAAATGAGATTAATTTAATATATAATAAAAGGTTGAAGGAATGTTGCTTCCACAAA	421
Dd	10858	TTAATGTTTATAAATAATATAGAAATATAGTCGAAAGAGATAGTTTTTGTGTTTGAAA	10917
Oy	422	CACCTGTCACATTTTAACGTAGTTAGTTCAAGTAATGTATAGGTAATGTTCCATAA	481
Dd	10918	ATAAAGTTGAATGGTTATTTTGTAGGTAGATTTGAATTAATAGATTTTTTGG	10977
Oy	482	TGATTATAAAGAGAGATTTTCAAACATTTTACCATTAT	520
Dd	10978	TTTTTTTTTAAGTTTTTTTTTGTTTTTTTTTTATATATTTT	11016

RESULT 20  
US-10-311-455-1602  
Sequence 1602 Annot:carion tte/10011AEE

```

:
:
: PUBLICATION NO. US20030143606A1
:
: GENERAL INFORMATION:
:
: APPLICANT: OLEK, Alexander
:
: APPLICANT: PREPENBROCK, Christian
:

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: DE 10043826.1  
PRIOR FILING DATE: 2000-09-01

```

/ OUT: 12:00:00.000000000
/ SEQ ID NO: 1602
/ LENGTH: 12138
/ TYPE: DNA
/

```

```

FEATURE:.....
;
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1602

```

Query Match	7.7%	Score	48.2	DB	12	Length	12138
Best Local Similarity	46.1%	Pred. No.	2.4				
Matches	161	Conservative	0	Mismatches	188	Indels	0
				Gaps			0

QY 162 TTTTGAACAATTTCTACCTATTTTACAGATATTTCGGTTGAGGTTACGCACA 221

Db 3113 TTTTATTTATTTTATTAATGATTTTGTTTATTAATTTTATTAATTAGATTGCA 3172

.QY		222	GTCAGATCATGCTTTTATTATGTGCACAGAGTGACTTTTTATGGCAAT	281
Dδ		3173	AATAAATGATTAAATGTTTAATTTTATATATAATTAATTTATTTAT	3232

Qy 282 ATTACCTTTTATGCGTTACGGCATGATTCATTCACGTACAAAGTGGTGCAATTATG 341  
Db 3233 ATTAAGTGTTAATTTAATGCAAAATGTAATTTAGCTTTGTTATATGTGACAATTTTAAAT 3292

QY 342 GACTCAAAAGTTTAAATTAATGAGTAATTTATATATATAAAGGTTGAAG 401  
DB 3293 GTTTATTAACGGTTTGTAGTATAAAGAGGTGTAATAAATGTTTATATAGTAATTAAGAAAG 3352  
QY 402 AATGTTTGCTTCAACAACACTTGTGACTTTTACGTAAGTTAGTCAAGTAATGT 461  
DB 3353 AATGTTTATATGTTTATATATTTTATTTATTTAGAGTATTTATATATTTATTTATATAT 3412  
QY 462 TATAGTATTTGTCATATATATATTAAGAAGAGGTTTCAACTTT 510  
DB 3413 TTTTATATTTGTTTATTTTATTTTATTTATTTAGTATGAGATTTTGAAGATT 3461  
RESULT 21  
US-10-311-455-1893  
; Sequence 1893, Application US/10311455  
; Publication No. US20030143606A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIERENBERG, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ  
; FILE REFERENCE: 5013.1014  
; CURRENT APPLICATION NUMBER: US/10/311,455  
; PRIOR FILING DATE: 2002-12-16  
; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 2424  
; SEQ ID NO 1893  
; LENGTH: 12590  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-1893  
Query Match 7.7%; Score 48.2; DB 12; Length 12590;  
Best Local Similarity 45.0%; Pred. No. 2.4; Indels 1; Gaps 1;  
Matches 220; Conservative 0; Mismatches 268; Indels 1; Gaps 1;  
QY 80 AATCAAGAGGTTACTAAAGCATGACTACTGACAGAGGATGCTGGGCTTGTACTAT 139  
DB 752 ATTAGTAGTTAAATTAATGAAATGATTTGTATAGTTATGAAGTTCGGAAGTTGA 811  
QY 140 ATTGAGCGTAAATGAGCATGCTATTTGAACAATTTCTACTATATTTTACAGATCAT 199  
DB 812 AATGAGTGTATATGGGTAATAAATTAAGAGAAATTTGTTTGTATTTTGAATTTAGG 871  
QY 200 TTCGTTGAGTACAGACAGTCAATGCTTATCTTATCTTATCTTATCTTATCTTATCT 258  
DB 872 GATGTTGTTGATTTTGTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 931  
QY 259 AATGCTATTTATGATGCAATATTTACTTTTATGCTGTTACGCGATGATTTCTATTA 318  
DB 932 GAATTTTGTGTTTATGTTTATATTTATTTTGTGTTGTTGTTGTTGTTGTTGTTGTTG 991  
QY 319 CGGTAAAGAGTGGGCAATATGAGTACAAAGGTTTAAATTTAAATGAGTAAATTT 378  
DB 992 TTTTATTTTGTAGAGAGATTAAGTTAGTTAATGTTTATTTTAAATTTTAAATTA 1051  
QY 379 AATATATATAAGGTTGAAGAGATGTTGCTTCAACAAACCTGTTGCACTTTTA 438  
DB 1052 TTTGTAATAAATTTTATTTTAAATTAAGTATTTATTTAGTTTGAAGAGTATTTT 1111  
QY 439 CGTATGTTAGTCAAGTAATTTGTTATAGTATTTGTTCAATATATGATTTTAAAGAGAG 498  
DB 1112 GGGGTTTGTATATTAATGTTTGAATCGTTATATTTTGTGTTTATTTTAAAGATTA 1171

QY 439 TTTTAACTTTTACATATATATGAGTCTTCAAGTGGAGATTAATGCTT 558  
DB 1112 TTTATTAAGTTTGTGTTTATTTTATTTTATTTTGAATATTTGATATGTTTAAATGTTGAA 1231  
QY 559 AATACAGGA 567  
DB 1232 TAGGTAGGA 1240  
RESULT 22  
US-10-311-455-950  
; Sequence 950, Application US/10311455  
; Publication No. US20030143606A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIERENBERG, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ  
; FILE REFERENCE: 5013.1014  
; CURRENT APPLICATION NUMBER: US/10/311,455  
; PRIOR FILING DATE: 2002-12-16  
; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 2424  
; SEQ ID NO 950  
; LENGTH: 18512  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-950  
Query Match 7.7%; Score 48.2; DB 12; Length 18512;  
Best Local Similarity 49.5%; Pred. No. 2.8; Indels 2; Gaps 1;  
Matches 152; Conservative 0; Mismatches 153; Indels 2; Gaps 1;  
QY 168 AACATTTTCTACTATATTTTACAGAAATCATTTGCTTGGAGTTACAGACAGTCACT 227  
DB 4706 AATATTATAGTATTTAGTTGAATGATTAATTTGTTTGAAGTATATATAATTTTGT 4765  
QY 228 ATATCTTATTTATTTATGTTGCTCAAGATAGCTATTTTATGATGCCAATTTACT 287  
DB 4766 TTTTATATTTGTTTATTTATTTTATTTTATGTTGTTATGTTATATTTTATATAT 4825  
QY 288 TTTTATGCTGTTAGGAGATGATTTCAATTCGTTACAGAGTGGCATTAAGACTAC 347  
DB 4826 GGTATTTGTTAATATAGAAATGATTTTATATAGTTGTTTATATTTATG--TGG 4883  
QY 348 AAAGCTTTTAAATTTAAATGAGATTAATTTATTAATAAAGGTTGAAGAATGTT 407  
DB 4884 GAAAGAAAGAGAGATTAATAAATAATTAATATATATGTTTATATTTATTTA 4943  
QY 408 TCGTTTCAACAACACTTTGCACTTTTACGTAAGTATGTTCAAGTAATTTGTTATAG 467  
DB 4944 TGTAGTTATTTTATTTAGAGTTTATTTATTTTATTTTATTTTATTTTAAATTTGTTA 5003  
QY 468 TATGTT 474  
DB 5004 TTTT 5010  
RESULT 23  
US-10-311-455-1900  
; Sequence 1900, Application US/10311455  
; Publication No. US20030143606A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander

APPLICANT: PIEPENROCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation  
FILE REFERENCE: 5013.1014  
CURRENT APPLICATION NUMBER: US/10/311.455  
PRIOR FILING DATE: 2002-12-16  
PRIOR APPLICATION NUMBER: PCT/EP01/07537  
PRIOR FILING DATE: 2001-07-02  
PRIOR APPLICATION NUMBER: DE 10032529.7  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: DE 10043826.1  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 2424  
SEQ ID NO 1900  
LENGTH: 13427  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-1900

Query Match 7.7%; Score 48; DB 12; Length 13427;  
Best Local Similarity 47.4%; Pred. No. 2.7;  
Matches 181; Conservative 0; Mismatches 195; Indels 6; Gaps 1;

Qy 234 TTTATTTATTTATGTTGCTCAAGAGATAGCTATTTTATGAGCAATATTACTTTTAT 293  
Db 10874 TTTATTTATTTATGTTGCTCAAGAGATAGCTATTTTATGAGCAATATTACTTTTAT 10933  
Qy 294 TGTGTTACGCGATGATTTGATTAACGTGACAAAGTGTGCTATTTAGACTACAAAGT 353  
Db 10934 ATTAATTTTATGATTTGATTTTAAATTTAGTATTTTAAATTAATAGTATTTA 10993  
Qy 354 TTTT-----AAATTTAATGAGTAATTTAATTAATAAAGGCTTGAAGAGATCT 407  
Db 10994 TTTTATGATTAATTTGAAATAGTAATTAATTAATTAATTAATTAATTTTATTT 11053  
Qy 408 TGCTTCTCAACAACACTGTTGCTTGCATTTAGTATTTAGTCAAGTAATTTATTAAG 467  
Db 11054 TTTTATTAAGTAAGAAATTTGTTTATGATTTGTTGAATGATTAATTAAGAGGTG 11113  
Qy 468 TATGTTCCATATATGATTAATAAAGAGAGTTTCAACTTTTACCATTAATTAATGC 527  
Db 11114 AGTTTGTGTTTATTAAGATTAATTAATTTTGTAGATTTGATTTATGTTAAAA 11173  
Qy 528 AAGTCTTCAGGTGCGCATTAATGCTTAATACAGAAATGACTGTTTATTAATAC 587  
Db 11174 AAGTTTTTGGAGTAATTTATTTGTTGCGGAAATATATATTTAAATTAATAGT 11233  
Qy 588 GCTAATTCCTATGACAATTAAT 609  
Db 11234 GGTGTTGATTTAGTATTAATTT 11255

RESULT 24  
US-10-311-455-80  
Sequence 80, Application US/10311455  
Publication No. US20030143606A1  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENROCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation  
FILE REFERENCE: 5013.1014  
CURRENT APPLICATION NUMBER: US/10/311.455  
PRIOR FILING DATE: 2002-12-16  
PRIOR APPLICATION NUMBER: PCT/EP01/07537  
PRIOR FILING DATE: 2001-07-02  
PRIOR APPLICATION NUMBER: DE 10032529.7  
PRIOR FILING DATE: 2000-06-30  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 2424  
SEQ ID NO 80  
LENGTH: 19972  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-80

Query Match 7.7%; Score 47.8; DB 12; Length 19972;  
Best Local Similarity 46.4%; Pred. No. 3.5;  
Matches 199; Conservative 0; Mismatches 222; Indels 8; Gaps 1;

Qy 181 TATATTTTACAGATGATTTGCTTGGAGTTACACAGACGATATATGCTTATTT 240  
Db 691 TTTTATTTTATGAGAGAGATGTAATTAATTTGTAATTAATTTTATTTATTTT 750  
Qy 241 ATTATGTTGCTCAAGAGATAGCTATTTTATGAGCAATATTACTTTTATGCTGTT 300  
Db 751 ATATTAATTTGTTGAGATTAATTAATTAATTTATTTTATTTTATTTTATTTT 810  
Qy 301 AGGCGATGATTTGATTAACGTGACAAAGTGTGCTATTTAGACTACAAAGCTTTTAA 360  
Db 811 GAGATAGGCTTTATTTATTTAGTTTACGTTGAGGTGAGTGGCGTGAATTAATGTTT 868  
Qy 361 TTTAATAGAGTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 420  
Db 869 -----TGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 922  
Qy 421 ACACTTGTGCTTCAAGAGATAGCTATTTTATGAGCAATATTACTTTTATGCTGTT 480  
Db 923 GTTGATTTATGATTAACGTAATTAATTTAGCTTTGTTATTTTATTTTATTTTATTT 982  
Qy 481 ATGATTAATAAGAGAGCTTTTCAACTTTTATTAATTAATTAATTAATTAATTAAT 540  
Db 983 AGGTTTTTATGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1042  
Qy 541 GTGCGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 600  
Db 1043 TTTTGGATTTGATGTTGATTTATTTATTTATTTATTTATTTATTTATTTA 1102  
Qy 601 ACAATTAAT 609  
Db 1103 GTAAATTAAT 1111

RESULT 25  
US-10-311-455-1530  
Sequence 1530, Application US/10311455  
Publication No. US20030143606A1  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENROCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation  
FILE REFERENCE: 5013.1014  
CURRENT APPLICATION NUMBER: US/10/311.455  
PRIOR FILING DATE: 2002-12-16  
PRIOR APPLICATION NUMBER: PCT/EP01/07537  
PRIOR FILING DATE: 2001-07-02  
PRIOR APPLICATION NUMBER: DE 10032529.7  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: DE 10043826.1  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 2424  
SEQ ID NO 1530  
LENGTH: 7110  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)



US-10-311-455-1530

Query Match 7.7%; Score 47.6; DB 12; Length 7110;  
Best Local Similarity 46.3%; Pred. No. 2.6;  
Matches 237; Conservative 0; Mismatches 264; Indels 11; Gaps 2;

QY 80 AATCGAAGAGGTACTAAGATGACGACGACAGGAGGAGGCGGCTTGCATTT 139  
DB 585 AAGTGAAGGTTGTATTAAGGAGATATTAAGGTTTGGGATTAATGATGATG 5912  
QY 140 ATTCAGCGCTTAATGGAGCTCATTTTGAACAATTTCTACTATTTTACAGATCAT 199  
DB 5913 ATTAGTAACTATGATTAATTAATTAATCGTTTAAATAGTATTAATTAATTT 5972  
QY 200 TTCCGTTTAAAGTTACAGACAGTC-----AGTAACTGTTTATTTATATGTTGCT 252  
DB 5973 ATGTGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6032  
QY 253 CAAGAGATAGCTATTTTATTTGATGCAATTTACTTTTATTTGCTGTACGCAATGAT 312  
DB 6033 GAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6092  
QY 313 TCATTAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 372  
DB 6093 TTGTTAAGATGTTTGTGTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6152  
QY 373 AATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 432  
DB 6153 AGGTGATTTTAAAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6212  
QY 433 CTTTAACTAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 492  
DB 6213 ATTTTGTGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6268  
QY 493 GGAGAGTTTCAACTTTTACATTAATTAATTAATTAATTAATTAATTAATTAAT 552  
DB 6269 ATTATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6328  
QY 553 ATGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 584  
DB 6329 TTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6360

RESULT 26  
US-10-311-455-1486  
; Sequence 1486; Application US/10311455  
; Publication No. US20030143606A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ  
; FILE REFERENCE: 5013.1014  
; CURRENT APPLICATION NUMBER: US/10/311,455  
; CURRENT FILING DATE: 2002-12-16  
; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; NUMBER OF SEQ ID NOS: 2424  
; LENGTH: 11092  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-1486

Query Match 7.7%; Score 47.6; DB 12; Length 11092;  
Best Local Similarity 46.1%; Pred. No. 3.1;

Matches 201; Conservative 0; Mismatches 229; Indels 6; Gaps 1;

QY 162 TTTGAAACAATTTCTACTATTTTTCAGAAATCATTTGGTTGAGTTACAGACA 221  
DB 10390 TTTTGTGGAGTTTGTGTTTGGGTTTGTGTTTATTTGTTTATTTGTTGAATATT 10449  
QY 222 GTCAGATTAATGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 281  
DB 10450 GTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10509  
QY 282 ATTACTTTTATTTGCTGTACGCAATGATTTCTACTAGTGTACAGTGTGATTAAT 341  
DB 10510 TTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10569  
QY 342 GACTCAAAAGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 401  
DB 10570 TTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10629  
QY 402 AATGTTGCTTCTCAAC-----AACAAGTTGCACTTTTACGTTAGTCAAGT 455  
DB 10630 AATTTATAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10689  
QY 456 AATGTTATAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 515  
DB 10690 TGTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10749  
QY 516 ATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 575  
DB 10750 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10809  
QY 576 TGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 591  
DB 10810 TATTTTGTGGTTA 10825

RESULT 27  
US-10-311-455-999  
; Sequence 999; Application US/10311455  
; Publication No. US20030143606A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ  
; FILE REFERENCE: 5013.1014  
; CURRENT APPLICATION NUMBER: US/10/311,455  
; CURRENT FILING DATE: 2002-12-16  
; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; NUMBER OF SEQ ID NOS: 2424  
; LENGTH: 6056  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-999

Query Match 7.6%; Score 47.4; DB 12; Length 6056;  
Best Local Similarity 45.7%; Pred. No. 2.7;  
Matches 165; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

QY 161 ATTGGAACAATTTCTACTATTTTTCAGAAATCATTTGGTTGAGTTACAGAC 220  
DB 3425 AATGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3484  
QY 221 AGTCAGTATAGCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 280



Db 3924 AATATTTGATTTTATTTTGGTTTTTTTTTTTAAATTAATAAATTTTGTAGG 3983  
QY 550 TATATGCTTAATACAGAAATAGTACTGTTTATATACCTAATT 594  
Db 3984 TTGTTGTTTGTGTAAGTTGTTTGTGTTTAAAGATAATT 4028

RESULT 30  
US-10-239-676-57  
; Sequence 57, Application US/10239676  
; Publication No. US20030082609A1  
; GENERAL INFORMATION:  
; APPLICANT: PIERENBERCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
; FILE REFERENCE: 5013.1003  
; CURRENT APPLICATION NUMBER: US/10/239,676  
; CURRENT FILING DATE: 2002-09-24  
; PRIOR APPLICATION NUMBER: PCT/EP01/03968  
; DE 10019058.8  
; DE 10019173.8  
; DE 10032529.7  
; DE 10043826.1  
; PRIOR FILING DATE: 2001-04-06  
; 2000-04-06  
; 2000-04-07  
; 2000-06-30  
; 2000-09-01  
; NUMBER OF SEQ ID NOS: 228  
; SEQ ID NO 57  
; LENGTH: 6385  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
; NAME/KEY: unsure  
; LOCATION: (5467)  
; US-10-239-676-57

Query Match 7.6%; Score 47.4; DB 14; Length 6385;  
Best Local Similarity 46.7%; Pred. No. 2.8; Mismatches 246; Indels 2; Gaps 2;  
Matches 217; Conservative 0;

QY 131 TTGCTATTTATTCAGGCGTAATGAGAGCTCATTTGAAACAATTTCTAATATTTTAA 190  
Db 3565 TTTTATTTTATTTGAATTTTATTTTATTTGATTTATTTTATTTTAAATTTAAGTTT 3624  
QY 191 CAGAAATTCATTCGGTTGAGTTAGCAGACAGTCAATATGCTTATTTATTTATTTAGTTG 250  
Db 3625 AATTTTATTTTATTTAAATAGTTAAATAATTTTATTTAGATAGATTTTATTTTGTG 3684  
QY 251 CTC-A-AGAGATAGCTATTTTATTTGATGCAATATTTACTTTTATTTGCTGTACGGCAGT 309  
Db 3685 TTTAATATGTTTTTATTTATTTTGGGTTTAAAGCATTTTGTGTTTAAAGT 3744  
QY 310 ATTTCATTCAGTGTACAGTGTGTCATTATGACATCAAAAGTTTAAATTTAAATG 369  
Db 3745 TTAGGATTAATGCTGTGAATTAATATTTTATTTTAAATAGTTTATTTAATGAAAAAT 3804  
QY 370 AGTAATTTAATATAATAAAGGTTGAAGAGATGTTTCTTCCAAACAACCTGTT 429  
Db 3805 TTTAATATAATAAATAAAGAGATGAATTTATTTTATTTTGAATTAATATTTATTT 3864  
QY 430 CGACTTTTACGTAGTTAGTCAAGTAATGTTATAGATTTGTCATATATGATTAATA 489  
Db 3865 TTAATGATTA-GTTGTTTATTTAATATTTATTTTATTTTAAATATGTTAATTTAT 3923  
QY 490 AAAAGAGATTTTCAAACTTTTACATTAATATGATCAAGTCTTCAAGGTGTGAGAT 549  
Db 3924 AATATTTGATTTTATTTTGTGTTTTTTTTTTTAAATTAATAAATTTTGTAGG 3983

QY 550 TATATGCTTAATACAGAAATAGTACTGTTTATATACCTAATT 594  
Db 3984 TTGTTGTTTGTGTAAGTTGTTTGTGTTTAAAGATAATT 4028

RESULT 31  
US-10-311-455-1822  
; Sequence 1822, Application US/10311455  
; Publication No. US20030143606A1  
; GENERAL INFORMATION:  
; APPLICANT: PIERENBERCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determination of Cytosine Methylation  
; FILE REFERENCE: 5013.1014  
; CURRENT APPLICATION NUMBER: US/10/311,455  
; CURRENT FILING DATE: 2002-12-16  
; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 2424  
; SEQ ID NO 1822  
; LENGTH: 7057  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
; US-10-311-455-1822

Query Match 7.6%; Score 47.4; DB 12; Length 7057;  
Best Local Similarity 50.4%; Pred. No. 2.9; Mismatches 161; Indels 6; Gaps 2;  
Matches 170; Conservative 0;

QY 167 AACAATTTTCTACATATTTTAAAGAAATCATTTGCTTGAAGTTACAGACAGTAC 226  
Db 3051 AATATATTTTATTTTGAATTTTGTGAATTTATTTATTTAAGATAGATTTATATAA 3110  
QY 227 TATATGCTTTATTTTATTTATGTTCTCAAGATAGTATTTTATTTATGCAATATTTAC 286  
Db 3111 TTAATGTTAATTTTATTTATTTATGTTCTAATAATTAATAATAATATTAATAAATA 3170  
QY 287 TTTTATGCTGTACGCGATGATTTCAATTCATTCATTCATTCATTCATTCATTCATTCAT 346  
Db 3171 TTTTAAATTTTGAATTAATGAATTAATTTATTTAAGTGTGATTAATATTTTGTG 3226  
QY 347 CAAGCTTTTAAATTTAAATGAGTAAATTAATAATAAAGGTTGAAGAATGT 406  
Db 3227 -AAGGCTTTTGTATTTATTTATTTAGATTTTATTTAAATTAAGTATGTTGAAATTT 3285  
QY 407 TTGCTTCAACAAACACTTG-TTCGACTTTTACGTATTTAGTTCAAGTAATTTGTTATA 465  
Db 3286 TTAATTTAATTTAATTTATTTGTTTGAATTAATGAATTAATTAATTTGTTGATTTTATA 3345  
QY 466 GGTATTTTCCATTAATGATTAATAAAGAGAGATTT 502  
Db 3346 AATATTTTGTGAGATTAATAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 3382

RESULT 32  
US-10-240-485-148  
; Sequence 148, Application US/10240485  
; Publication No. US20030148327A1  
; GENERAL INFORMATION:  
; APPLICANT: PIERENBERCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Metastasis

FILE REFERENCE: 5013.1007  
CURRENT APPLICATION NUMBER: US/10/240.485  
CURRENT FILING DATE: 2002-10-02  
PRIORITY APPLICATION NUMBER: PCT/EP01/03970  
PRIORITY FILING DATE: 2001-04-06  
PRIORITY APPLICATION NUMBER: DE 10019058.8  
PRIORITY FILING DATE: 2000-04-06  
PRIORITY APPLICATION NUMBER: DE 10019173.8  
PRIORITY FILING DATE: 2000-04-07  
PRIORITY APPLICATION NUMBER: DE 10032529.7  
PRIORITY FILING DATE: 2000-06-30  
PRIORITY APPLICATION NUMBER: DE 10043826.1  
PRIORITY FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 202  
SEQ ID NO 148  
LENGTH: 7057  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-240-485-148

Query Match  
Best Local Similarity 50.4%; Score 47.4; DB 12; Length 7057;  
Pred. No. 2.9; Mismatches 161; Indels 6; Gaps 2;

167 AACCAATTTCTACTATATTTTACAGATCATTTGGTTGAGTTACAGACAGTCAG 226  
1051 AAAAATTTTATTTTGAATTTTGGAAATTTTGAATTTTGAATTTTGAATTTTGA 3110  
227 TATATGCTTATTTATTTATTTATGCTTCAAGAGATGCTTATTTATGATGCAATATTAC 286  
3111 TAAATGTTAATTTTATTTATTTATGTTGTTAAATTTAAATTTAAATTTAAATTTAA 3170  
287 TTTTATGCTGTTACGAGCAGATGATTCATTCAGTACAGAGTGGTCATATGAGACTA 346  
3171 TTTTAAATTTTGAATTTGAATTTGATTTGAGATTTGATTTGATTTGATTTGATTTG 3226  
347 CAATGTTTAAATTTAAATTTGAGTAAATTTATTTATTTAAATTTGTTGAAGCAATGT 406  
3227 -AAGGTTTGTGTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3285  
407 TTGCTTCTCAACAAACCTTG-TTCGACTTTTACGATTTAGTTCAAGTAAATTTTATA 465  
3286 TTATTTATTTATTTAAATTTGTTGTTAGAAATTTATGATTTAAATTTGTTGATGATTTTATA 3345  
466 GGAATTTGCTCATATATGATTTATTTAAAGAGAGATTTT 502  
3346 AAATGTTTGAAGATTTAAATTTAGGAAACGTAAATTT 3382

## RESULT 33

US-10-311-455-1702  
Sequence 1702, Application US/10311455  
Publication No. US20030143606A1  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENBROCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OR INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect  
FILE REFERENCE: 5013.1014  
CURRENT APPLICATION NUMBER: US/10/311.455  
CURRENT FILING DATE: 2002-12-16  
PRIORITY APPLICATION NUMBER: PCT/EP01/07537  
PRIORITY FILING DATE: 2001-07-02  
PRIORITY APPLICATION NUMBER: DE 10032529.7  
PRIORITY FILING DATE: 2000-06-30  
PRIORITY APPLICATION NUMBER: DE 10043826.1  
PRIORITY FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 2424  
SEQ ID NO 1702  
LENGTH: 17721

TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-1702

Query Match  
Best Local Similarity 46.5%; Score 47.2; DB 12; Length 17721;  
Pred. No. 4.5; Mismatches 187; Conservative 0; Mismatches 213; Indels 2; Gaps 1;

125 TGGGCGTTCGTTATTTATTTACGCGCTATATGGACGCTCATTTTGAACAATTTCTACTATA 184  
11680 TGGTTTTTTTAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 11739  
185 TTTTACAGATCATTTGCGTTGAGGTTACAGCA--CAGTCGATATATGCTTATTTAT 242  
11740 TTTTATTTAAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 11799  
243 TTAATGCTCAAGATGATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 302  
11800 TATTTTATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 11859  
303 GGCATGATTTTCATTTACGCTGACAGTTGGTGCATTTAGACTACAAAGTTTAAATT 362  
11860 AGAATGTAAGATTTAAGAAATTTTAAATTTTATTTATTTATTTATTTATTTATTTATTTACG 11919  
363 TAAATGAGTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 422  
11920 TATTTATGATGTTGTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 11979  
423 ACTTGTGCTTCAAGATTTAGTTTCAAGTAAATTTATTTATTTATTTATTTATTTATTTATTTATTT 482  
11980 GGTAAATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 12039  
483 GATTTATTTAAAGAGATTTTCAACTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 524  
12040 GAATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 12081

## RESULT 34

US-10-311-455-1424  
Sequence 1424, Application US/10311455  
Publication No. US20030143606A1  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENBROCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OR INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect  
FILE REFERENCE: 5013.1014  
CURRENT APPLICATION NUMBER: US/10/311.455  
CURRENT FILING DATE: 2002-12-16  
PRIORITY APPLICATION NUMBER: PCT/EP01/07537  
PRIORITY FILING DATE: 2001-07-02  
PRIORITY APPLICATION NUMBER: DE 10032529.7  
PRIORITY FILING DATE: 2000-06-30  
PRIORITY APPLICATION NUMBER: DE 10043826.1  
PRIORITY FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 2424  
SEQ ID NO 1424  
LENGTH: 19787  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-1424

Query Match  
Best Local Similarity 50.0%; Score 47.2; DB 12; Length 19787;  
Pred. No. 4.7; Mismatches 143; Indels 12; Gaps 1;  
Matches 155; Conservative 0; Mismatches 143; Indels 12; Gaps 1;  
162 TTTTGAACAATTTCTACTATATTTTACAGATCATTTGGTTGAGTTACAGACACA 221



Db 3517 GGGTTAGGTATAGTGGTTTATGTTATATTTAGTAT 3554

RESULT 37  
US-10-240-453-269  
; Sequence 269, Application US/10240453  
; Publication No. US20030148326A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA  
; TITLE OF INVENTION: Transcription  
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated  
; FILE REFERENCE: 5013.1009  
; CURRENT APPLICATION NUMBER: US/10/240,453  
; CURRENT FILING DATE: 2002-10-02  
; PRIOR APPLICATION NUMBER: PCT/EP01/03973  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: DE 10019058.8  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019173.8  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 350  
; SEQ ID NO 269  
; LENGTH: 15743  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-240-453-269

Query Match 7.5%; Score 46.8; DB 12; Length 15743;  
Best Local Similarity 46.5%; Pred. No. 5.2;  
Matches 185; Conservative 0; Mismatches 212; Indels 1; Gaps 1;

QY 129 GCTTCTATTATTTATTCAGCGGTAATGGAGCTCATTTTGAACAATTTTCTACATATTTT 188  
DB 12578 GTTTTATAGGTTTAAATGATTTTGTGTTAGTTTGTGTTAGTTGGATTTATAGATA 12637

QY 189 TAGGAAATCATTTGGTTTGAGTTACGACAGTCGATATGCTTTATTTATTTATGT 248  
DB 12638 TGTATTTATATTTAGTTAATTTTGTGTTATTTTGTAGTAGAGTTGGGTTTTTTAAGT 12697

QY 249 TGCTCAAGATAGTCTATTTTATGATGCCAATATTACTTTTATGCTGTACGGCATG 308  
DB 12698 TGGTATAGTGTGTTTAAATTTTCGATTTTGGGATTTATTTGTTGGTTTTTAAAG 12757

QY 309 -GATTTCAATACGTGTACAGTTGGTGCATTTAGACTACAAAGTTTAAATTTAAAT 367  
DB 12758 TGTGGAATATATATGATGATTAATGGCTTGGAGAGATGGGTTTTTATTTGAT 12817

QY 368 GGAATTAATTTAATATATAAAGGTTGAAGAAGATTTGCTTCTCAACAACACTTG 427  
DB 12818 TAGAAGTATTTGGTGTGATTAAGAGTGTGAGTGAAGGAATTTGTTGTTGTTT 12877

QY 428 TTTCACTTTTCAGTATTTAGTTCAAGTAATTTGTAATGTAATTTGTCATATATGATTA 487  
DB 12878 TTTTATATGTTAATATATAGTTTAAAGGATATATTTAGATTTGATAGGGTTGTTATTT 12937

QY 488 TAAAGAGAGCTTTCAAACTTTTACCATATATAT 525  
DB 12938 ATTATGATTTTATTTTAAATTTGTAAGATATATATAT 12975

RESULT 38  
US-10-311-455-1593  
; Sequence 1593, Application US/10311455

Publication No. US20030143606A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ  
; TITLE OF INVENTION: Cytosine methylation  
; FILE REFERENCE: 5013.1014  
; CURRENT APPLICATION NUMBER: US/10/311,455  
; CURRENT FILING DATE: 2002-12-16  
; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 2424  
; SEQ ID NO 1593  
; LENGTH: 6254  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-1593

Query Match 7.5%; Score 46.6; DB 12; Length 6254;  
Best Local Similarity 46.0%; Pred. No. 4;  
Matches 157; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

QY 162 TTTTGAACAATTTTCTACATATATTTTACAGATCATTTTCGTTGAGTTACAGACA 221  
DB 627 TTTTACATATTTTAAATTTATTTTAAATATATAGTGGAGTATTTTAAAAATA 686

QY 222 GTGATATATGCTTATTTATTTATTTGTTGCTCAAGATATGCTATTTATTTATGATCCAT 281  
DB 687 TAAATTTAGATTTTATGATTTATTTTGTGTTAAAGTTTGAATTTTATTTATTTT 746

QY 282 ATTACTTTTATTTGCTGTACGGCATGATTTTCATTCATGTCAGTGTGTCATTATG 341  
DB 747 AGAATTTATTTAAATTTTATTTATTTATTTGATTTGATTTGTTTAAATATTA 806

QY 342 GACTACAAGGTTTAAATTTAAATGAGTAATTTAATATTAATAAAGGTTGAAAG 401  
DB 807 GGATGAAGAAGTATTTTAAATATTTTAAATTTGTTTAAATTTGTTTAAAGAAATGTGAAAT 866

QY 402 AATGTTGCTCTCAACAACACTTGTTCGACTTTTACGATTTAGTTAGTTCAAGTAATGT 461  
DB 867 TAGTTTATTTTAAATTTTGTGTAAGAGTGTGTTGTTATTTTAAATTTATGAAATAT 926

QY 462 TATAGTATTTGTCATATATGATTATTAAGAAGAGATTTT 502  
DB 927 ATTTGGTAAATATATTTAAAGTATGTTTAAATATATATATAT 967

RESULT 39  
US-10-204-708-11  
; Sequence 11, Application US/10204708  
; Publication No. US20030141852A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
; TITLE OF INVENTION: by Assessing DNA Methylation  
; FILE REFERENCE: 5013.1012  
; CURRENT APPLICATION NUMBER: US/10/204,708  
; CURRENT FILING DATE: 2003-05-06  
; PRIOR APPLICATION NUMBER: PCT/EP01/03971  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: DE 10019058.8  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019173.8  
; PRIOR FILING DATE: 2000-04-07





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2003, 02:48:32 ; Search time 1935 Seconds  
(without alignments)  
7812.602 Million cell updates/sec

Title: US-10-010-160-1

Sequence: 1 atgtctgatgcccaagtaa.....aatattgcagtcgcagatc 622

Scoring table: IDENTITY\_NUC

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : EST:4

1:	em	estba:	*
2:	em	esthu:	*
3:	em	estin:	*
4:	em	estum:	*
5:	em	estov:	*
6:	em	estpl:	*
7:	em	estro:	*
8:	em	hrc:	*
9:	gb	estl:	*
10:	gb	est2:	*
11:	gb	hrc:	*
12:	gb	est3:	*
13:	gb	est4:	*
14:	gb	est5:	*
15:	em	estfun:	*
16:	em	estom:	*
17:	em	gs_hum:	*
18:	em	gs_inv:	*
19:	em	gs_pln:	*
20:	em	gs_vrt:	*
21:	em	gs_fun:	*
22:	em	gs_mam:	*
23:	em	gs_mus:	*
24:	em	gs_pro:	*
25:	em	gs_rod:	*
26:	em	gs_pbg:	*
27:	em	gs_vr1:	*
28:	gb	gsml:	*
29:	gb	gs62:	*

**Pred. No.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	60.4	9.7	996	29	CNS000FTH	AL071063 Drosophila
C 2	60.2	9.7	1101	29	CNS000KES	AL077453 Drosophila
C 3	59.2	9.5	1154	13	BX436282	BX436282 BX436282
C 4	59.2	9.5	1200	13	BX415878	BX415878 BX415878

C	5	58.6	9.4	886	28	BH177277	008	L	22	
C	6	58.5	9.4	866	29	CNS07JUX				AL614225 T3 end of
C	7	57.8	9.3	1201	13	BX360615				BX360615
C	8	57.8	9.3	1201	13	BK24465				BK24465
C	9	57.4	9.2	568	13	BU497346				BU497346 PESToab6
C	10	57.4	9.2	1201	13	BK439779				BK439779 BX439779
C	11	56.8	9.1	563	13	BU495402				BU495402 PESToab7
C	12	56.8	9.1	1101	29	CNS017V2				AL108536 Drosophill
C	13	56.6	9.1	734	29	CNS010MP				AL099163 Drosophill
C	14	56.6	9.1	1101	29	CNS00088				AL063632 Drosophill
C	15	56.6	9.1	1201	13	BX461824				BX461824
C	16	56.2	9.0	1201	13	AL565455				AL565455 AL565455
C	17	56	9.0	1146	13	BX462546				BX462546 BX462546
C	18	55.8	9.0	1101	29	CNS003DO				AL066580 Drosophill
C	19	55.4	8.9	977	13	BK415819				BK415819 BX415819
C	20	55.2	8.9	473	13	BQ452226				BQ452226 PESToaa9
C	21	54.8	8.8	945	29	CNS04DOK				AL285149 Telraedon
C	22	54.8	8.8	1201	9	AL536104				AL536104 AL536104
C	23	54.6	8.8	1101	29	CNS00EVL				AL069706 Drosophill
C	24	54.4	8.7	994	13	BK414650				BK414650 BX414650
C	25	54.4	8.7	1092	29	CNS020K7				AL175696 Telraedon
C	26	54.2	8.7	1021	29	CNS014Y7				AL104032 Drosophill
C	27	53.8	8.6	854	9	AL522840				AL522840 AL522840
C	28	53.8	8.6	1076	13	BX338020				BX338020 BX338020
C	29	53.6	8.6	1201	29	CNS0167M				AL106339 Drosophill
C	30	53.6	8.6	1101	29	CNS014XJ				AL104737 Drosophill
C	31	53.4	8.6	1887	13	BK441520				BK441520 BX441520
C	32	53.2	8.6	1201	13	BK347585				BK347585 BX347585
C	33	52.8	8.5	1757	29	BX214654				BX214654 Danilo rer
C	34	52.8	8.5	1056	13	BK415058				AL410689 BX415058
C	35	52.8	8.5	1101	29	CNS016L1				AL106896 Drosophill
C	36	52.6	8.5	999	13	BK380865				BK380865 BX380865
C	37	52.6	8.5	1200	29	CNS016C0				AL106578 Drosophill
C	38	52.4	8.4	759	29	CNS06QXV				AL411257 T7 end of
C	39	52.4	8.4	1101	29	CNS00E07				AL069440 Drosophill
C	40	52	8.4	928	29	CNS00DXY				AL071865 Drosophill
C	41	52	8.4	1094	29	CNS012E2				AL101513 Drosophill
C	42	52	8.4	1177	13	BK399460				BK399460 BX399460
C	43	52	8.4	1201	13	BK381083				BK381083 BX381083
C	44	52	8.4	1201	13	BK426629				BK426629 BX426629
C	45	51.8	8.3	1045	13	BX456814				BX456814 BX456814

## ALIGNMENTS

RESULT 1  
CNS00FTUH/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

CNS00FTUH 996 bp DNA linear GSS 03-JUN-1999  
Drosophila melanogaster genome survey sequence T1T3 end of BAC;  
BACR31021 of RFL1-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
AL071063  
AL071063.1 GI:4951105  
GSS.  
Drosophila melanogaster (fruit fly)  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 996)  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Ooeegawa and  
Aron Mammosier in Pletier de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

source

Location/Qualifiers

1..996

/organism="Drosophila melanogaster"

/mol\_type="genomic DNA"

/db\_xref="taxon:7227"

/clone="BACR31021"

/clone\_11b="RPCI-98"

/note="end : TET3"

BASE COUNT 383 a 164 c 81 g 171 t 197 others

ORIGIN

Query Match 9.7%; Score 60.4; DB 29; Length 996;

Best Local Similarity 34.5%; Pred. No. 0.1; Mismatches 223; Indels 0; Gaps 0;

Matches 154; Conservative 69; Mismatches 223; Indels 0; Gaps 0;

164 TTGAACAATTTCTACATATTTTACAGATTCGTTGAGTTACAGACAGT 223

962 TTWMAAAATWMAATWMAATTTTATTTTBTBSSSSSSSSSSSSGRRAAA 903

224 CAGTATATGCTTATTTATTTATGTTCTCAAGATGATGATTTATGATGCAAT 283

902 AAAAAAAMWMAATTTTWTWMAATTTTATTTATTTATTTATTTATTTT 843

284 TACTTTTATGCTGTTAGCGATGATTTTCAATGATGATGATGATGATG 343

842 TTTTATTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTT 783

344 CTACAAGGTTTAAATTTTAAATGAGTAAATTTTAAATTTTAAAGGTTGAAGAA 403

782 AATWMAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 723

404 TGTTCCTTCTCAACAACACTGTTGACCTTTTACGATGATGATGATGATG 463

722 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 663

464 TAGGATTTGCTCATATGATTTAAAGAGAGTTTCAACTTTTACCATTAAT 523

662 AACAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 603

524 ATGCAAGTCTTCAAGTGTGAGATTTATGCTTATATACAGATGATGATGAT 583

602 AATWTTTATTTTAAATATKADTWMAATTTTATTTTATTTTATTTTATTT 543

584 ATACGCTAATTCCTATGACATTAAT 609

542 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 517

RESULT 2

CNS00KB5 1101 bp DNA 11near GSS 03-JUN-1999

LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC:

DEFINITION BAC39P05 of RPCI-98 library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

ACCESSION AL077453 GI:4956930

VERSION AL077453.1

KEYWORDS GSS.

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)

AUTHORS Genoscope.

TITLE Direct Submission

## JOURNAL

## COMMENT

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuo Osewaga and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

source

Location/Qualifiers

1..1101

/organism="Drosophila melanogaster"

/mol\_type="genomic DNA"

/db\_xref="taxon:7227"

/clone="BACR39P05"

/clone\_11b="RPCI-98"

/note="end : T7"

BASE COUNT 540 a 0 c 139 g 252 t 170 others

ORIGIN

Query Match 9.7%; Score 60.2; DB 29; Length 1101;

Best Local Similarity 42.3%; Pred. No. 0.11; Mismatches 185; Conservative 10; Mismatches 242; Indels 0; Gaps 0;

113 TTTCTATATTTTACAGATTCATTCGTTGAGTTACAGACAGTCAATGAT 232

578 TTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 519

233 CTATATTTATTTATGCTCAAGATGATGATTTTATGATGCAATATTTAT 292

518 TTTTAAAAAATTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 459

233 TTGCTGTTACGCGATTCATTTCAATGCTGTCACAGTTGTCATATGAC 352

458 TTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 399

333 TTTTAAATTTAAATGAGTAAATTTATTTAAATGAGTGAAGAAATGTTGCT 412

398 TTTTAAAAATTTAAAAATTTAAAAATTTAAATTTNTTNTTNTTNTTNTT 339

413 CTCAACAACACTGTTGACCTTTACGTTAGTTCAAGTATTTATGATATG 472

338 ATTNNNNNNNTTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 279

473 TTCCATATATGATTTAAAGAGAGTTTCAACTTTTACCATTAATTTATG 532

278 TTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 219

533 CTTCAGGTGTGACATTTATGCTTATATACAGATGATGATGTTTATATG 592

218 WTTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 159

593 TTCCATATGATTTAT 609

158 TTTATTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 142

RESULT 3

BX436282 1124 bp mRNA 11near EST 15-MAY-2003

LOCUS BX436282 Homo sapiens THYMUS Homo sapiens cDNA CSOCAP001YC01

DEFINITION 5-PRIME, mRNA sequence.

ACCESSION BX436282

VERSION BX436282.1 GI:30787521

**KEYWORDS** EST.  
**SOURCE** Homo sapiens (human)  
**ORGANISM** Homo sapiens  
**REFERENCE** 1 (bases 1 to 1124)  
**AUTHORS** Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi; L1,W.B., Gruber,C., Jessee,J. and Polayes,D.  
**TITLE** Full-length cDNA libraries and normalization  
**JOURNAL** Unpublished  
**COMMENT** Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. Contact : Feng Liang Email : filiang@lifetech.com URL :  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0CAP001AB010P1.  
**FEATURES**  
 source  
 1. .1124  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0CAP001YC01"  
 /issue\_type="THYMUS"  
 /clone\_lib="Homo sapiens THYMUS"  
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed  
 with a NotI-oligo(dt) primer. Five prime end enriched  
 double-strand cDNA was digested with Not I and cloned into  
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
 Library was not normalized."  
 BASE COUNT 129 a 29 c 109 g 435 t 422 others  
 ORIGIN  
 Query Match 9.5%; Score 59.2; DB 13; Length 1124;  
 Best Local Similarity 23.8%; Pred. No. 0.17; Mismatches 189; Indels 0; Gaps 0;  
 Matches 102; Conservative 139; Mismatches 189; Indels 0; Gaps 0;  
 Oy 144 AGCGTATGAGGAGCATTTTGAACAATTTCTACATATTTTACAGATCATTTG 203  
 Db 694 RRRRRRRRRDTTTTTTTDRDTTTTTTTDDTTTTTTTDRRRRRDTTTTT 753  
 Oy 204 GTTGAAGTTACAGCAGCAGTATATGCTTATTTATGTTGCTCAAGATGAC 263  
 Db 754 TTTTDDKTTTTRRRRTTTRRRRRRTTTRRRRTTDTTTTTRRRRRRRRR 813  
 Oy 264 TATTTATGATGCAATATTACTTTTATGCTTACGCAATGATTCATTAAGT 323  
 Db 814 RRTTTTTTTTTTTRRRRRRRRRRTTTTTTTTTTTTTTTTDDRRRRRTT 873  
 Oy 324 ACAAGTTGTCATATGATGACAAAGTTTAAATTAATGAAGATTAATTAAT 383  
 Db 874 TDDRRRRRRRRDTTTTTTTDDTTTTTTTDDDDTTTDRRRRRRRRRRRR 933  
 Oy 384 AATAAAGGTTGAAGATGTTGCTTCAACAACACTGTCGACTTTTACGTAG 443  
 Db 934 RRRRRRTTDRRRRTTTTTTTTTTRRRRRDDBDKRRRTTMDTTTTTTTT 993  
 Oy 444 TTTGTTCAATATGTTATAGTATGCTTCAATATGATTTAAAGAGAGTTTC 503  
 Db 994 TTRRRRRRRRTTTRRRRRRRRRRRRTTTRRRRTTDRRRRRRTT 1053  
 Oy 504 AAATCTTTTACATATATGCAAGTCTTCAGTGAGATTAATGCTTAATAC 563  
 Db 1054 DRRRRRTTTRRRRTTDDTTTTTTTTTTTDRRRRRDTTTTTTTTTTTT 1113  
 Oy 564 AGAATAGT 572  
 Db 1114 RRRRRRTT 1122

RESULT 4  
 EX415878/c 1200 bp mRNA linear EST 15-MAY-2003  
 LOCUS BX415878

**DEFINITION** BX415878 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP008Y104  
**ACCESSION** 5-PRIME, mRNA sequence.  
**VERSION** BX415878  
**KEYWORDS** EST.  
**SOURCE** Homo sapiens (human)  
**ORGANISM** Homo sapiens  
**REFERENCE** 1 (bases 1 to 1200)  
**AUTHORS** L1,W.B., Gruber,C., Jessee,J. and Polayes,D.  
**TITLE** Full-length cDNA libraries and normalization  
**JOURNAL** Unpublished  
**COMMENT** Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. Contact : Feng Liang Email : filiang@lifetech.com URL :  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0CAP008BE02QPL.  
**FEATURES**  
 source  
 1. .1200  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0CAP008Y104"  
 /issue\_type="THYMUS"  
 /clone\_lib="Homo sapiens THYMUS"  
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed  
 with a NotI-oligo(dt) primer. Five prime end enriched,  
 double-strand cDNA was digested with Not I and cloned into  
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
 Library was not normalized."  
 BASE COUNT 428 a 70 c 85 g 447 t 170 others  
 ORIGIN  
 Query Match 9.5%; Score 59.2; DB 13; Length 1200;  
 Best Local Similarity 37.2%; Pred. No. 0.17; Mismatches 207; Indels 1; Gaps 1;  
 Matches 158; Conservative 59; Mismatches 207; Indels 1; Gaps 1;  
 Oy 162 TTTGAACAATTTTCTACTATATTTTACAGATCATTTGCTTGAAGTACGACA 221  
 Db 1081 WTTAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1022  
 Oy 222 GTCAATATAGCTTATTTATTTATGTTGCTCAAGATATGCTATTTATGATCA 281  
 Db 1021 AAATAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 962  
 Oy 282 ATTACTTTTATGCTGTTACGAGATTTCTTAAGTCAAGTGGCATTAATG 341  
 Db 961 ATTTTTTTTATTTATTTATTTTAAATTAATTAATTAATTAATTAATTA 902  
 Oy 342 GACTACAAAGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 401  
 Db 901 AAATTTTSTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 842  
 Oy 402 AATGTTGCTTCAACAACAACCTGTTGCACTTTTACGATTTAGTCAATATGT 461  
 Db 841 WTTTTTAAATTAATTAATTAATTAATTAATTTTAAATTTTAAATTTTAA 782  
 Oy 462 TATAGTAT-TGTCATATATGATTAATAAGAGAGTTTCAACTTTTACCATTA 520  
 Db 781 TAAATTAATTTTATGATTAATTAATTAATTAATTAATTAATTTCTTTAA 722  
 Oy 521 ATTATGCAAGTCTTCAAGTGTGAGATTAATGCTTAATTAACGAATAGTCTG 580  
 Db 721 WAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 662  
 Oy 581 TATAT 585  
 Db 661 YMAAT 657

	RESULT 5	
	BH177277/c	
	LOCUS	
DEFINITION	BH177277	886 bp DNA linear GSS 19-OCT-2001
	008_L_22-rev SmbAC1 Schistosoma mansoni genomic clone 008L22 5',	
	genomic survey sequence.	
ACCESSION	BH177277	
VERSION	BH177277.1	
KEYWORDS	GI:16275873	
SOURCE	GSS.	
ORGANISM	Schistosoma mansoni	
	Schistosoma mansoni	
	Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;	
	Stigeleidia; Schistosomatoidae; Schistosomidae; Schistosoma.	
REFERENCE	1 (bases 1 to 886)	
AUTHORS	Le Paslier,M.-C., Pierce,R.J., Merlín,F., Hirai,H., Wu,W., Williams	
	,D.L., Johnston,D., Loyerre,P.T. and Le Paslier,D.	
TITLE	Construction and characterization of a Schistosoma mansoni	
	bacterial artificial chromosome library	
	Genomics 65 (2), 87-94 (2000)	
JOURNAL	20247247	
MEDLINE	10783255	
PUBMED	Other GSSs: 008_L_22-21	
COMMENT	Contact: Pierre RJ	
	INSERM U 167	
	Institut Pasteur de Lille	
	1 rue du Professeur A. Calmette, 59019-Lille, France	
	Tel.: (33) (0)3 20877783	
	Fax: (33) (0)3 20877888	
	Email: Raymond.Pierce@pasteur-lille.fr	
	CNS sequencing ID=DG0AA008DFIIBP1	
	Plate: 008 row: L column: 22	
	Seq primer: M13 reverse primer	
	Class: BAC ends	
	High quality sequence stop: 886.	

FEATURES	SOURCE	LOCATION/Qualifiers
	1..886	Location/Qualifiers
		/organism="Schistosoma mansoni"
		/mol_type="genomic DNA"
		/strain="Puerto-Rican"
		/db_xref="taxon:6183"
		/clone="008122"
		/sex="mixed"
		/dev_stage="cercariae"
		/lab_host="Biomphalaria glabrata"
		/clone_1b="SmbAC1"
		/note="Vector: pBeloBAC 11; Site 1: Hind III; Partially Hind III digested and size-selected S. mansoni cercarial DNA was ligated into Hind III digested pBeloBAC 11 vector and used to transform E. coli DH10. The complete library contains 23808 clones from 4 independent sizing-ligation-transformations. Average insert size ranges from 70-127 Kb and genome coverage is 7.9-fold."
BASE COUNT	376 a	22 c 49 g 288 t 151 others
ORIGIN		

[illegible]

```

D6 660 TTTTAAATTAATTAATAAAAMTTTTTAAAAATATTTTTTTTTTAAAAAATTAAATTWA 601
QY 413 CTCAACAAACACTGTTCGACCTTTTACCTAGTTTAGTCAAGTAATGTTATAGTATG 472
D6 600 TWAAMTTAAATTAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 541
QY 473 TTCCATATATGATTTATAAAAGAGAGTTTCAAACTTTTACATTAATATTATGCAAGTC 532
D6 540 TTTTAATTAATTAATTTKTATAAAAAAAKMAATTTTKKTWTMDKKKARAT - KKKAKKAKTTKT 483
QY 533 CTTCAGAGTGGCAGATTATATGCTTAATACAGAAATAGTCTGTTTATATACGTTAA 592
D6 482 TKKRAWAATKTTTATATATAATKKAANAATATATAATTTTAAATBIAATAMSSTRA 423
QY 593 TTCCATATGCAATTT 606
D6 422 TKTAATTTTSTTTT 409

```

RESULT 6	
CNS07JUX/c	
LOCUS	CNS07JUX 886 bp DNA linear GSS 30-NOV-2001
DEFINITION	T3 end of clone 008DF1 of library SmbAcl from strain Puerto-Rican
	of <i>Schistosoma mansoni</i> , genomic survey sequence.
ACCESSION	AF614235
VERSION	AF614235.1 GI:16027459

KEYWORDS	GSS.
SOURCE	Schistosoma mansoni
ORGANISM	Schistosoma mansoni

REFERENCE  
1 (bases 1 to 886)  
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;  
Strigidae; Schistosomatoidea; Schistosomatidae; Schistosoma.

**AUTHORS** DE PASILLER, M. C., PIERCE, R. C. J., WELSH, J. R., HILL, R. L., MU'W, M. J., WILLIAMS, D. L., JOHNSTON, D., LOVERDE, P. T., and DE PASILLER, D.

**TITLE** Construction and characterization of a *Schistosoma mansoni* bacterial artificial chromosome library

**JOURNAL** Genomics 65 (2), 87-94 (2000)

MEDLINE	20247247
PUBMED	10783255
REFERENCE	2 (bases 1 to 886)

Genoscope - Centre National de Séquençage :  
 Submitted (05-OCT-2001) Direct Submission  
 JOURNAL TITLE Genoscope.

COMMENT  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
BP 191 91006 EVRY cedex - FRANCE (E-mail : [segrete@genoscope.cns.fr](mailto:segrete@genoscope.cns.fr))  
Partially Hind III digested and size-selected S. mansoni cercarial

**FEATURES**  
DNA was ligated into Hind III digested pBeloBAC 11 vector and used to transform *E. coli* DH10. The complete library contains 23808 clones from 4 independent sizing-ligation-transformations. Averaging insert size ranges from 70-127 kb and genome coverage is 7.9-fold.  
location/Qualifiers

```
source
1. .886
/organism="Schistosoma mansoni"
/mol_type="genomic DNA"
"Data Dir"
```

```
BASE COUNT      376 a      22 c      49 g      288 t      151 other
ORIGIN
```

	Query Match Similarity	9.4%	Score 58.6;	DB 29;	length 886;	
	Best Local Similarity	40.6%	Pred. No. 0.22:			
	Matches	176;	Conservative	50;	Mismatches 199;	Indels 9; Gaps 2
Oy	173	TTTTTACATATTTTTTCAGAAATCATTTCCGTTTGAGGTTCACAGCAGTCAGTAATAG	232			
		: : : : :				
Dd	833	TTTTTTTTTTTWTAAAAAATTTTATATTTTTTTTTTTTTTAAATTAATAATAMAMTTT	774			
Oy	233	CTTATATTATTATGTCCTCAGAAGATAGCAATTTTATATGAATGCCAATATTACTTTTAA	292			
	: : : :	: : : :	:	:	:	:
Dd	773	TATTMTTWTTTTAATTTWATAAAAATTAATATATTAMTTTTTTTTTAAAAATTTTATTTT	714			

QY 293 TTGCTGTCGCGATGATTTTCATTACGTCGTCAGAGTTGTGCTATGACTACAAAG 352  
 Db 713 TAAATTTTAAATTTTAAAMAAA-----AAAAATTTTAAATTTTAAATTTAAATTT 661  
 QY 353 TTTTAAATTTTAAATGAGTAATTTAATATATATAAAGGTTGAAGAAATGTTGTT 412  
 Db 660 TTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 601  
 QY 413 CTCACAAACACTGTCGATTTTACGATTTTACGATTTTACGATTTTACGATTTTACG 472  
 Db 600 TTAATTAATTAATTTAAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTT 541  
 QY 473 TTCCATATATGATTAATAAGAGAGAGTTTCAACTTTTACATATATATGCAAGTC 532  
 Db 540 TTTTAAATTTAAATTTTAAAMAAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 483  
 QY 533 CTTCAGGTGTCGAGATTTATGCTTAATACAGAAATGTTCTGTTTATATACGCTAA 592  
 Db 482 TKTATMMATKTTTTAAATTAATTTAAATTTAAATTTAATTTAATTTAATTTAATTT 423  
 QY 593 TTCCATGACAAAT 606  
 Db 422 TKTATTTTSTTTT 409

RESULT 7  
 EX360615/c 1201 bp mRNA linear EST 05-MAY-2003  
 LOCUS BX360615 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA  
 DEFINITION clone CSOD1072Y014 5-PRIME, mRNA sequence.  
 ACCESSION BX360615  
 VERSION BX360615.1 GI:30372450  
 KEYWORDS EST.

SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1201)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CSOD1072BH07Qp1.

FEATURES  
 SOURCE location/Qualifiers  
 1..1201

/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CSOD1072Y014"  
 /issue\_type="PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo (dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."  
 BASE COUNT 386 a 141 c 149 g 254 t 271 others  
 ORIGIN

Query Match 9.3%; Score 57.8; DB 13; Length 1201;  
 Best Local Similarity 31.8%; Pred. No. 0.32;  
 Matches 159; Conservative 90; Mismatches 251; Indels 0; Gaps 0;  
 QY 92 TTACTAAAGCATGACTAGCAGGAGGCTGGGCTTGCTATTTATTCAGCGTAA 151  
 Db 1190 WMAAAMAAAMWMTTMMADWMAAAMADDTTKKTKDDKDADAAMWMAAAMWT 1131

QY 152 TGGAGCTCATTTTGAACAAATTTTCTACTATATTTTACGAATCATTTGGTTGAGG 211  
 Db 1130 TTTTMMWMTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1071  
 QY 212 TTACAGCAGCTAGCTATGCTTATTTATTTATTTGCTCAAGAGTACGATTTAT 271  
 Db 1070 TTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1011  
 QY 272 TGATGCAATATATCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 331  
 Db 1010 TTAATTAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 951  
 QY 332 GTGCAATATGAGCTACAAAGGTTTAAATTTAAATTTAAATTTAAATTTAAATTT 391  
 Db 950 ATAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 891  
 QY 392 GGTGAAGAAGATTTGCTTCTCAACAAACACTTTGCGACTTTTACGTTAGTTC 451  
 Db 890 TTTTDTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 831  
 QY 452 AAGTAATGTTATAGTATTTGCTCATATATGATTTAAAGAGTTTCAAACTTTT 511  
 Db 830 AATAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 771  
 QY 512 TACCATTAATTTATGCAAGCTTTCAGGTGTCAGATTTATATGCTTAATACAGAAAT 571  
 Db 770 TDAATATKTATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 711  
 QY 572 TACTGTTTATATACGCTA 591  
 Db 710 TMTWMTTMMKTKTKTKMM 691

RESULT 8  
 EX424465/c 1201 bp mRNA linear EST 13-MAY-2003  
 LOCUS BX424465 Homo sapiens PLACENTA Homo sapiens CDNA clone CSODE002YP24  
 DEFINITION 5-PRIME, mRNA sequence.  
 ACCESSION BX424465  
 VERSION BX424465.1 GI:30659708  
 KEYWORDS EST.

SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1201)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 4172.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS5A004ZD12Qp1&cluster=4172.r. Contact :  
 Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS5A004ZD12Qp1.

FEATURES  
 SOURCE location/Qualifiers  
 1..1201

/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CSODE002YP24"  
 /issue\_type="PLACENTA"  
 /note="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched,  
 double-strand cDNA was digested with Not I and EcoRV sites of the pCMVSPORT 6 vector.  
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.

BASE COUNT 508 a 103 c 175 g 319 t 96 others  
 ORIGIN

Query Match  
 Best Local Similarity 38.0%; Score 57.8; DB 13; Length 1201;  
 Matches 167; Conservative 43; Mismatches 230; Indels 0; Gaps 0;

131 TTGCTATTATTCAGGCGTAATGGAGCTCATTTTGAACAATTTCTCTATTTTAA 190  
 1106 TTTTATTTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1047  
 191 CAGAAATCATTTGGTTGAGTTAGCAGACGATGATATGCTTATTTATTTAGTTG 250  
 1046 TTTATTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 987  
 251 CTCAGATGATGCTATTTTATTTATTTGATGCAATTTACTTTTATTTGCTTACG 310  
 986 TTTTAAATTTTATTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 927  
 311 TTTCAATTCAGTTCAGATGTTGGTGCATTTAGCAATCAAGTTTAAATTTAAATG 370  
 926 ATATATATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 867  
 371 GTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 430  
 866 TTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 807  
 431 GACTTTTATTTAGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 490  
 806 ATATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 747  
 491 AAGGAGCTTTTCAACTTTTATTTATTTATTTATTTATTTATTTATTTATTTAT 550  
 746 ATATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 687  
 551 ATATGCTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 570  
 686 ATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 667

RESULT 9  
 BU497346/c 568 bp mRNA linear EST 11-SEP-2002  
 LOCUS PF5870ab62b05.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium  
 DEFINITION BU497346  
 ACCESSION BU497346  
 VERSION BU497346.1 GI:22793540  
 KEYWORDS EST.  
 SOURCE Plasmodium falciparum 3D7  
 ORGANISM Plasmodium falciparum 3D7  
 REFERENCE 1 (bases 1 to 568)  
 AUTHORS Fukuoka, Y., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D.,  
 Tang, K., Cole, R., Martin, J., Wylie, T., Dante, M., Theising, B.,  
 Marra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B.,  
 Bowers, Y., Gibbons, M., Ritey, E., Bennett, J., Jentes, E., Ronko, T.,  
 Tsagarelis, V., Richey, J., Wadkins, J., Kennedy, S., Levinson, D.,  
 Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinson, D.,  
 Waterston, R., Wilson, R., and Sibley, D.  
 TITLE Unpublished  
 JOURNAL  
 COMMENT WashU Plasmodium EST Project  
 Contact: L. David Sibley  
 WashU Plasmodium EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu  
 Library was constructed by Debopam Chakrabarti DNA sequencing by:  
 Washington University Genome Sequencing Center for information on  
 obtaining a clone please contact: L. David Sibley  
 (sibley@brcim.wustl.edu), Washington University  
 Seg primer: -40UP from Gibco

FEATURES  
 source  
 High quality sequence stop: 425.  
 Location/Qualifiers  
 1..568  
 /organism="Plasmodium falciparum 3D7"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:36329"  
 /lab\_host="DH10B (GeneHog, Invitrogen, Inc.)"  
 /clone\_idb="Plasmodium falciparum 3D7 asexual cDNA"  
 /note="Vector: pBluescript SK plus, Site 1: EcoRI, Site 2:  
 XhoI; Library was constructed by Debopam Chakrabarti.  
 Total RNA samples were isolated from mixed stage  
 asexuals (0.1%) lysed P. falciparum 3D7 infected  
 erythrocytes by the acidic guanidinium-phenol chloroform  
 method. The poly A+ RNA was isolated by the polyAT-Tract  
 mRNA isolation system (Promega, WI) using streptavidin  
 Magneticbead particles. Directional cDNA libraries were  
 constructed by oligo d(T) priming of poly(A)+ RNA (5mg)  
 into EcoRI and XhoI sites of 1 ZapII vector using the Zap  
 cDNA synthesis kit (Stratagene, CA). The average size of the  
 cDNA inserts in the library was between 1.0 and 1.5kb.  
 Clones were mass excised using the Exasist helper phage  
 (Stratagene), the phagemids were precipitated with PEG  
 8000 and extracted with phenol/chloroform. Phagemid DNA  
 was electroporated into DH10B cells."

BASE COUNT 295 a 64 c 57 g 152 t  
 ORIGIN

Query Match  
 Best Local Similarity 46.2%; Score 57.4; DB 13; Length 568;  
 Matches 190; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

199 TTTGGTTGAGTTAGCAGACGATGATATGCTTATTTATTTATTTATTTATTTATTT 258  
 568 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 509  
 259 ATAGCTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 318  
 508 TTGGATTAACAGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 449  
 319 CGTGTACAGTTTGTGCTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 378  
 448 CTTTAAAGTGTGTTTCTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 389  
 379 AATATATTAAGGTTGAAAGGATTTGCTTCAACAACACTTTGTCAGCTTTA 438  
 388 TTATTCATACAGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 329  
 439 CGTATTTAGTTCAAGTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 498  
 328 ATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 269  
 499 TTTTCAACTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 558  
 268 TCTGAAGATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 209  
 559 AATACAGATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 609  
 208 TTATTTAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 158

RESULT 10  
 BX439779/c 1201 bp mRNA linear EST 15-MAY-2003  
 LOCUS BX439779 Homo sapiens PLACENTA Homo sapiens cDNA clone CS080141F05  
 DEFINITION 3-PRIME, mRNA sequence.  
 ACCESSION BX439779  
 VERSION BX439779.1 GI:30771778  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1201)

AUTHORS Li, W.B., Gruber, C., Jeessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 COMMENT

Genoscope - Centre National de Sequencage  
 Contact: Genoscope  
 BP 191 91006 Evry cedex - France  
 Email: seqre@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 3370.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0DE014CC03NP1&cluster=3370.r. Contact :  
 Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0DE014CC03NP1.  
 Location/Qualifiers

FEATURES  
 source  
 1..1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DE014FE05"  
 /issue\_type="PLACENTA"  
 /clone\_lib="Homo sapiens PLACENTA"  
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed  
 with a NotI-oligo (dT) primer. Five prime end enriched,  
 double-strand cDNA was digested with Not I and cloned into  
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
 Library was not normalized."  
 BASE COUNT 455 a 150 c 90 g 289 t 207 others  
 ORIGIN

Query Match 9.2%; Score 57.4; DB 13; Length 1201;  
 Best Local Similarity 37.1%; Pred. No. 0.38;  
 Matches 151; Conservative 60; Mismatches 192; Indels 4; Gaps 1;  
 Oy 162 TTTTGAACAATTTCTACTATATTTTACGAATCATTCGGTTGAGTTACAGACA 221  
 Db 1107 TTTTGAATTTTAAATTTTCTACTATATTTTACGAATCATTCGGTTGAGTTACAGACA 221  
 Oy 222 GTGAGTAAAGCTTTATTTATTTATTTGTCCTCAAGAGATGACTATTTATTTGAGCCAT 281  
 Db 1047 ATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 988  
 Oy 282 ATTTCTTTTATTTGCTGATGAGATGATTTCACTAGTACAGATGAGTGGTCAATTATG 341  
 Db 987 TTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 928  
 Oy 342 GACTACAAAGCTTTTAAATTTAAATGAGTAAATTTATTAATAAAGGTTGAAAG 401  
 Db 927 AATTATATATTTTAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 868  
 Oy 402 AATGTTGCTTCGCA-----CAACACTGTTGCACTTTTACGATTTAGTTGCAAGTAA 457  
 Db 867 TTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 808  
 Oy 458 TTGTTATAGATTTGCTCATATATGATTAATAAAGAGAGTTTCAACTTTTACCAT 517  
 Db 807 TTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 748  
 Oy 518 TATATTAAGCAAGTCTTCAGGTGCGAGATTAATGCTTAATACA 564  
 Db 747 TTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 701

RESULT 11  
 BU495402/c 563 bp mRNA linear EST 11-SEP-2002  
 LOCUS BU495402  
 DEFINITION falciparum 3D7 CDNA 5', mRNA sequence.  
 ACCESSION BU495402  
 VERSION BU495402.1 GI:22791596  
 KEYWORDS EST.  
 SOURCE Plasmodium falciparum 3D7

ORGANISM Plasmodium falciparum 3D7  
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 REFERENCE 1 (bases 1 to 563)  
 AUTHORS

Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D.,  
 Marra, M., Hillier, L., Martin, J., Wylie, T., Darte, M., Theising, B.,  
 Bowers, Y., Gibbons, M., Rilter, E., Bennett, J., Jentes, E., Ronko, I.,  
 Tagarelskikh, R., Belajgorod, L., Franklin, C., Carr, L., Grow, A.,  
 Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D.,  
 Mesteron, R., Wilson, R. and Sibley, D.  
 Washu Plasmodium EST Project

Unpublished  
 TITLE  
 JOURNAL  
 COMMENT  
 Contact: L. David Sibley  
 Washu Plasmodium EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu  
 Library was constructed by Debopam Chakrabarti DNA sequencing by:  
 Washington University Genome Sequencing Center For information on  
 obtaining a clone please contact: L. David Sibley  
 (sibley@borcim.wustl.edu), Washington University  
 Seq primer: -400P from Gibco  
 High quality sequence stop: 420.  
 Location/Qualifiers

FEATURES  
 source  
 1..563  
 /organism="Plasmodium falciparum 3D7"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:36329"  
 /lab\_host="DH10B (GeneHog, Invitrogen, Inc.)"  
 /clone\_lib="Plasmodium falciparum 3D7 asexual cDNA"  
 /note="Vector: Bluescript SK plus; Site 1: EcoRI; Site 2:  
 XhoI; Library was constructed by Debopam Chakrabarti.  
 Total RNA samples were isolated from mixed stage  
 saponin(0.1%-lysed P. falciparum 3D7 infected  
 erythrocytes by the acidic guanidium-phenol chloroform  
 method. The poly A+ RNA was isolated by the polyAT-Tract  
 mRNA isolation system (Promega, WI) using streptavidin  
 Magnosphere particles. Directional cDNA libraries were  
 constructed by oligo d(T) priming of poly(A)+ RNA (5mg)  
 into EcoRI and XhoI sites of 1 ZapII vector using the Zap  
 cDNA synthesis kit (Stratagene, CA). The average size of the  
 cDNA inserts in the library was between 1.0 and 1.5kb.  
 Clones were mass excised using the ExAssist helper phage  
 (Stratagene), the phagemids were precipitated with PEG  
 8000 and extracted with phenol/chloroform. Phagemid DNA  
 was electroporated into DH10B cells."

BASE COUNT 290 a 64 c 57 g 152 t  
 ORIGIN  
 Query Match 9.1%; Score 56.8; DB 13; Length 563;  
 Best Local Similarity 46.6%; Pred. No. 0.5; Indels 0; Gaps 0;  
 Matches 181; Conservative 0; Mismatches 207; Indels 0; Gaps 0;  
 Oy 222 GTGAGTAAAGCTTTATTTATTTATTTGTCCTCAAGAGATGACTATTTATTTGAGCCAT 281  
 Db 545 GTGATTTGATTTTGTGTTTCTTCTTCTACAGCAATTTTGGATTAATAGTTTGTGATT 486  
 Oy 282 ATTTCTTTTATTTGCTGTTAAGGATGATTTCACTAGTACAGATGAGTGGTCAATTATG 341  
 Db 485 TATATTTTAAATTTGTTATCTTTTGTCTCAGTACTTTTAAAGTGTGTTTCTTATT 426  
 Oy 342 GACTACAAAGCTTTTAAATTTAAAGAGATTAATTAATAAAGGTTGGAAG 401  
 Db 425 AATTATTTGTTTATTTAGTTGATACCTATTAATTTGTTATTAATAAAGATTTACTATA 366  
 Oy 402 AATGTTGCTTCGCAACAACACTGTTGCACTTTTACGATTTAGTTAGTTCAAGTATTTGT 461  
 Db 365 ATGACAGATTTGCTCAATTTATTTGTTGAATTAATTTATTTTAAATTTACTTCT 306  
 Oy 462 TATAGTATTTGCTCATATATGATTAATAAAGAGAGTTTCAACTTTTACCATATA 521  
 Db 305 TATGAGTATTTGTTCACTGATCATGATGATGAGTGTGGAAGAAATTATTAATATCATC 246



QY 522 TTATGCAAGTCTTTCAGGTGCGAGATTATAGCTTAATACAGAAATGATCTGTTT 581  
DB 245 ATCTTAAATATATTTCTTTCTTTCATTAATATTTGGTATTAAATATTAATTAATA 186  
QY 582 ATATAGCGTAATTCCTATGACATTAAT 609  
DB 185 ATTCTTATCATTTTATTAATTAATTAAT 158

RESULT 12  
LOCUS CNS017V2  
DEFINITION Drosophila melanogaster genome survey sequence Sp6 end of BAC  
BACN37J10 of DrosBAC library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION AL108536  
VERSION AL108536.1 GI:5628840  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster (fruit fly)  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1101)  
REFERENCE Genoscope.  
AUTHORS Direct Submission  
TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
COMMENT Determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelobAC11.

FEATURES  
source Location/Qualifiers  
1..1101  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACN37J10"  
/clone\_1lb="DrosBAC"  
/plasmid="pBelobAC11"  
/note="end : SP6"

BASE COUNT 268 a 174 c 114 g 366 t 179 others

ORIGIN  
Query Match 9.1%; Score 56.8; DB 29; Length 1101;  
Best Local Similarity 38.6%; Pred. No. 0.5;  
Matches 136; Conservative 43; Mismatches 173; Indels 0; Gaps 0;

QY 173 TTTTCACATATTTTACAGATCATTCGGTTGAGTACAGACAGTACATATG 232  
DB 750 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 809  
QY 233 CTTTATTTTATTTATGTTGCTCAAGATAGCTATTTATTTATGACCAATTTACTTTTA 292  
DB 810 WTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 869  
QY 293 TTGCTGTACGCGATTTTCATTAAGTGTACAGATGCGTGCATTATGACTACAAAG 352  
DB 870 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTA 929  
QY 353 TTTTAAATTTAAATGAGTAAATTTAATATAATAAAGGTGAAGAAGCTTGCT 412  
DB 930 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 989  
QY 413 CTCACAAACACTGTGTCAGCTTTTACGATTTAGTTCACAGTAATTTAGTATG 472  
DB 990 WTTTATTAATDATTWTTTAAATTTTATTTTATTTTATTTTATTTTATTTT 1049

QY 473 TTCCATATATATTTAATAAGAGATTTCAACTTTTACATTAATTA 524  
DB 1050 TTTTATTAATAATAATATTTTATTTATTTATTTATTTATTTATTTATTTA 1101

RESULT 13  
LOCUS CNS010MP/c  
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC  
BACN04L20 of DrosBAC library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION AL099163  
VERSION AL099163.1 GI:5610774  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster (fruit fly)  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 734)  
REFERENCE Genoscope.  
AUTHORS Direct Submission  
TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
COMMENT Determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelobAC11.

FEATURES  
source Location/Qualifiers  
1..734  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACN04L20"  
/clone\_1lb="DrosBAC"  
/plasmid="pBelobAC11"  
/note="end : T7"

BASE COUNT 288 a 62 c 2 g 211 t 171 others

ORIGIN  
Query Match 9.1%; Score 56.6; DB 29; Length 734;  
Best Local Similarity 35.5%; Pred. No. 0.55;  
Matches 124; Conservative 65; Mismatches 159; Indels 1; Gaps 1;

QY 178 TACTATATTTTACGAATCATTCGGTTGAGTTCAGACAGCTAGTATAGCTTTA 237  
DB 371 TTTTATGATGATTTTATTTTATTTATTTATTTATTTTATTTTATTTTATTT 312  
QY 238 TTTATTTATGTTGCTCAAGATAGCTATTTATTTATGATGCAATTTACTTTTATGCT 297  
DB 311 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 252  
QY 298 GTTACGAGATGATTTTCATTAAGTGTACAGTGTGTCATTATGACTACAAAGGTTT 357  
DB 251 WTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 192  
QY 358 AAATTTAAATGAGTAAATTTAATATAATAAAGGTGAAGAAG-AGTTTGCCTTCA 416  
DB 191 WAAAAAATTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTT 132  
QY 417 ACAACACTGTGTCAGCTTTTACGATTTAGTTCAGTAATTTATAGTATGTTCC 476  
DB 131 AAATTAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 72  
QY 477 ATATATGATTTAATAAGAGAGTTTCAACTTTTACATTATATTA 525  
DB 71 ATATAKATATATATATATTTTATTTTATTTTATTTTATTTTATTTTATTTT 23

RESULT 14

CNS000B8

### DEFINITION

Accession	Length	Type	Source
CNS00008	1101 bp	DNA	linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR01A24 of RPI1-8 library from Drosophila melanogaster (fruit fly), genomic survey sequence.			

ACCESSION

VERSION

## KEYWORDS

## SOURCE

## ORGANIS

## REFERENCE

## AUTHORS

TITLE

JOURNAL

**COMMENT**

Submitted (2-2-JUN-1999) Genoscope - Centre National de Séquençage  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazuhiro Osoegawa and  
Aaron Mammocci at the Picot de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCL-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2<sup>+</sup> cn bw sp. The same strain used for the BDGP's  
P1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

**FEATURES**  
**source**

Location/Qualifiers  
1. .1101

BASE COUNT	228 a	114 c	110 g	512 t	137 others
ORIGIN					

Query Match	9.1%;	Score 56.6;	DB 29;	Length 1101;
Best Local Similarity	38.3%;	Pred. No. 0.54;		
Matches 159;	Conservative 53;	Mismatches 199;	Indels 4;	Gaps 1;

Qy	167	AAACAATTTTCTACATATTTTTCACAGATCATTTTCGGTTTGAGGTTTCACAGACAGCTAG	226
Db	682	AAATATTTTTTCTKGTMTDITTTTGGTWTATTTTGGDAATWTATTTTTTWTATATMTM	741
Qy	227	TAAATGCTTATTTATTTATGTTGCTCAAGAGATAGCTATTTATGTATGCCAATATTA	286
Db	742	TWTWTAAATTTTATTTATTTTAAATTTTAAATTTTCTKGAATTTTWTATTTADTTTTTWTAT	801
Qy	287	TTTTTATGCTGTACCGCATGATTTTCATTAACGTACAAGTGGTCATTAATGAGCTA	346
Db	802	TTTWTATTTTWTATATMTMTTMTTMTTMTTMTTATTTATTTTMTATTTTMTTMTATTA	861
Qy	347	CAAAAGTTTTTAAATTTAAATGAGTAATTTTAAATATATAAAGGTTGAAAGGAATGT	406
Db	862	AAAAAAATTAATTTTAAATMT-----TAAATTTTAAATTTTAAATTTTAAATTAATATWT	917
Qy	407	TTGGCTTCAACAACAACCTGTGTGACCTTTACAGATGTTGTCACAGATTTGTTATAG	466
Db	918	TWATMTTAAATATTTATTTTMTTTTTTTTTTTTTTTTTTTTTTWTATATATATTTTA	977
Qy	467	GTATTTGTTCCATPATATGATTTATAAAGAGAGTTTCAACTTTTTTACCATTATATATAG	526

## RESULT 15

BX461824/

### DEFINITION

**ACCESSION**

**VERSION**

## KEYWORDS

## SOURCE

## REFERENCE

## AUTHORS

THE  
LITTLE  
TOWN

JOURNAL  
COMMITTEE

COMMENT:

Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: [sege@genoscope.cns.fr](mailto:sege@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
was not normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
8170.r For more information about this cluster, see  
<http://www.genoscope.cns.fr/>  
[cgibin.cluster.cgi?seq=CSDF034BA04QP1&cluster=8170.r](http://cgibin.cluster.cgi?seq=CSDF034BA04QP1&cluster=8170.r). Contact :  
Feng Liang Email : [liang@lifetech.com](mailto:liang@lifetech.com) URL :  
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600  
Paradise Avenue Genoscope Sequence ID : CSDF034BA04QP1.

**FEATURES**

**SOURCE**

Location/Qualifiers  
1. .1201

BASE COUNT	385 a	157 c	102 g	312 t	245 others
------------	-------	-------	-------	-------	------------

Query Match	9.1%;	Score 56.6;	DB 13;	Length 120;
Best Local Similarity	31.7%;	Pred. No. 0.54;		

[illegible]



Best Local Similarity 35.8%; Pred. No. 0.71;  
Matches 163; Conservative 67; Mismatches 223; Indels 2; Gaps 1;

```

QY 162 TTTTGAACAATTTCTACATATTTTACAGATTCATTCGTTGAGTTACAGCA 221
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 645 TTTTATATATTTATTTATTTTATTTTATTTTATTTTATTTTATTTATTT 704
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 222 CTCAGTATATGCTTATTTATTTATTTATTTGCTCAAGATGCTATTTT -ATGATGCCA 279
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 705 TTTATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 764
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 280 ATATATCTTTTATGCTGATGCGCATGATTCATTCAGTTCAGTTCAGTCA 339
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 765 GTGGTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 824
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 340 TGCACTACAAAGTTTAAATTTAAATGAGTAAATTTAATTAATTAATTAAGGTTGAA 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 825 TTTATTTTAAAAATTTAAATTTAAATTTTATTTTATTTTATTTTATTTTATTT 884
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 400 GGAATGTTGCTTCACAAACACTTGTGACTTTTACGATTTAGTTCAAGTAA 459
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 885 AAGKTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 944
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 460 GTTATGATTTGCTTCATATATGATTTTAAAGAGAGTTTCAACTTTTACATTA 519
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 945 KTKTAAATTTTATTTTATTTTAAARKMAAMDAAATTTTAAATTTTAAATTTT 1004
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 520 TATATGCAAGTCTTCAGTGTGAGTATATATGCTTATTAATGAGTATGCTGT 579
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1005 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1064
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 580 TTAATATAGCTTAATTCCTATGACATTTATTCAGT 614
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1065 TTTTATTTTAAATTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTT 1099
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

## RESULT 18

CNS003DQ

LOCUS 1101 bp DNA linear GSS 03-JUN-1999  
DEFINITION Drosophila melanogaster genome survey sequence TEXT end of BAC #  
BAC080109 of RPCI-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

ACCESSION

AL064580  
AL064580.1 GI:4941932

VERSION

KEYWORDS

SOURCE

ORGANISM

Drosophila melanogaster (fruit fly)  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Peerygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazuo Oseogawa and  
Aaron Mammoler in Peter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's  
P1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES

SOURCE

1. 1101

/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BAC080109"  
/clone\_1db="RPCI-98"  
/note="end : TEXT"

Query Match 9.0%; Score 55.8; DB 29; Length 1101;  
Best Local Similarity 41.6%; Pred. No. 0.77; Indels 3; Gaps 1;  
Matches 185; Conservative 23; Mismatches 234; Indels 3; Gaps 1;

```

QY 131 TTGCTATTTATTCAGGCTAATGAGTCAATTTGAAACATTTCTACTATATTTTA 190
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 18 TTTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 77
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 191 CAGATCATTTGCTGTTGAGGTTACACACAGTCAGTATATGCTTATTTATGTTG 250
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 78 TTTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTA 137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 251 CTCAGAGATGCTATTTTATGATGCCAATATCTTTTATGCTTACGCAATGCA 310
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 138 CGAAAAATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 311 TTTTATTCGTCACAGTGTGCTATTCATTCAGTCAAAAGTTTAAATTTAAATGCA 370
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 198 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 371 GTAATTTATATATATATATATATATATATATATATATATATATATATAT 430
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 256 -TTTWTGTAATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 314
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 431 GACTTATACGATTTGTTGTTCAATATGTTATATGATATGTTCCATATATGATTA 490
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 315 AATATATATATATATATATATATATATATATATATATATATATATATATAT 374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 491 AAGGAGATTTTCAACTTTTATTCATTTATTTATGCAAGTCTTCAGTGTGACAGT 550
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 375 WATATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 434
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 551 ATATGCTTAATATCAGATATGACT 575
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 435 WTTTATTTTAAATTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTT 459
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

## RESULT 19

BX415819

LOCUS 977 bp mRNA linear EST 13-MAY-2003  
DEFINITION BX415819 Homo sapiens THYMUS Homo sapiens cDNA clone CSOCAP0087G19  
5-PRIME, mRNA sequence.

ACCESSION

BX415819  
BX415819.1 GI:30645987

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: [sequef@genoscope.cns.fr](mailto:sequef@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Library was constructed by Life Technologies, a division of  
Invitrogen. Contact : Feng Liang Email : [liang@life-tech.com](mailto:liang@life-tech.com) URL :  
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600  
Paradise Avenue genoscope sequence ID : CSOCAP008AD100P1.  
Location/Qualifiers  
1. 977  
/organism="Homo sapiens"

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSOCAP008Y19"
/tissue_type="THYMUS"
/clone_lib="Homo sapiens THYMUS"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

```

BASE COUNT 240 a 144 c 98 g 380 t 115 others

Query Match 8.9%; Score 55.4; DB 13; Length 977;

Best Local Similarity 37.0%; Pred. No. 0.93;

Matches 141; Conservative 51; Mismatches 189; Indels 0; Gaps 0;

```

QY 145 GGGCTAATGGAGCGTCATTTTGAACAATTTCTACTATTTTACAGAAATCATTTGG 204
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 355 KKGKGRGDGDDKDKKKKKAAMAAAAADDDKDKKTTTWTWTWTWTWTWTWTWT 414
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 205 TTGAGGTACAGACAGTCAGTATGCTTATTTATTTATGTCACAGATAGCT 264
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 415 TTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 474
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 265 ATTTTATGATGCAATATTTACTTTTATGCTGTACGCGCATGATTCATTCGTGA 324
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 475 TTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 534
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 325 CAAGTGTGTCATATGACATCAAAAGTTTAAATTAATGAGTAATTAATA 384
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 535 TATTTTGTGTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 594
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 385 ATAAAGGTTGAAGAAATGTTGCTCTCAACAACATGTCGACTTACGTAGT 444
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 595 WTCTTAAAAAATTTAAATTTTAAAGTAAATWTAAATTTTWTWTWTWTWTWT 654
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 445 TTAGTCAAGTAATGTTATAGTATGTCCTATATGATTTAAAGAGAGTTTCA 504
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 655 TMAAATTAACCTAAATTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 714
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 505 AACTTTTACCATTAATTAAT 525
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 715 AAWTTTTTAAAMAAGATYMW 735
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :

```

## RESULT 20

B0452226 473 bp mRNA linear EST 29-MAY-2002

LOCUS PEST0aa3908.y1 Plasmodium falciparum 3D7 asexual cDNA plasmodium  
DEFINITION falciparum 3D7 cDNA 5', mRNA sequence.

ACCESSION B0452226  
VERSION B0452226.1 GI:21255338

KEYWORDS EST.  
SOURCE Plasmodium falciparum 3D7  
ORGANISM Plasmodium falciparum 3D7

REFERENCE 1 (bases 1 to 473)  
Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Page, D.,  
Merritt, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Treising, B.,  
Bowers, Y., Gibbons, M., Ritzer, E., Bennett, J., Jentes, E., Ronko, I.,  
Tagareishvili, R., Belaygorod, L., Franklin, C., Carr, L., Grow, A.,  
Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D.,  
Waterston, R., Wilson, R., and Sibley, D.

TITLE Unpublished  
JOURNAL Washu Plasmodium EST Project

COMMENT Contact: L. David Sibley  
Washu Plasmodium EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu

library was constructed by Debopam Chakrabarti. DNA sequencing by:  
Washington University Genome Sequencing Center For information on  
obtaining a clone please contact: L. David Sibley  
(sibley@orcim.wustl.edu), Washington University  
Seq primer: -400P from Gibco  
High quality sequence stop: 354.  
location/Qualifiers  
1. 473

## FEATURES

SOURCE

```

/organism="Plasmodium falciparum 3D7"
/mol_type="mRNA"
/db_xref="taxon:36329"
/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
/notes="Vector: pBluescript SK plus; Site 1: EcoRI; Site 2:  
XhoI; Library was constructed by Debopam Chakrabarti.  
Total RNA samples were isolated from mixed stage  
saponin(0.1%) lysed P. falciparum 3D7 infected  
erythrocytes by the acidic guanidinium-phenol chloroform  
method. The poly A+ RNA was isolated by the polyAT-tract  
mRNA isolation system (Promega, WI) using streptavidin  
magnosphere particles. Directional cDNA libraries were  
constructed by oligo d(T) priming of poly(A)+ RNA (5mg)  
into EcoRI and XhoI sites of 1 ZapII vector using the Zap  
cDNA synthesis kit (Stratagene, CA). The average size of the  
cDNA inserts in the library was between 1.0 and 1.5kb.  
Clones were mass excised using the ExAssist helper phage  
(Stratagene), the phagelids were precipitated with PBG  
8000 and extracted into phenol/chloroform. Phagemid DNA  
was electroporated into DH10B cells."

```

BASE COUNT 129 a 48 c 52 g 244 t

Query Match 8.9%; Score 55.2; DB 13; Length 473;

Best Local Similarity 46.4%; Pred. No. 1;

Matches 180; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

```

QY 222 GTCAATATATGCTTATTTATTTATGTTCTCAAGATACCTATTTATGATGCCAT 281
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2 GTCATTTGAATTTTGTGCTGTTTCTACAGCTAATTTTGATTAACAGTTTGTGCTT 61
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 282 ATTACTTTTATGCTGTTAGCGCATGATTTCAATGCTGTAACAGTTGGCATATG 341
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 62 TATATTTTGAATTTGTTATCTTTTGTCTCAGCTCTTTTAAAGTGTGTTCTTAT 121
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 342 GACTCAAAAGTTTAAATTTAAATGAGTAATTTAAATTAATTAATTAATTAATTA 401
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 122 ATTTAATTTGTTTATTTAGTTGATTAACCTATTTATTTATTAACAAGATTAATA 181
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 402 AATGTTGCTTCTCAACAACATGTTGCACTTTACGTAGTTAGTTCAAGTAATGT 461
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 182 ATGACTAGATGCTTAATTTATTTGTTGAATATTTTATTTTATTTTAAATTAATCT 241
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 462 TTATGATTTGTTCAATATGATTAATTAAGAGAGTTTCAACTTTTACCATTAATA 521
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 242 TATGAGTATGTTATCTGAATCAATGACAGTAGTCTGAAGAATTAATTTATTCATC 301
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 522 TATGCAAGTCTTCAAGTGTGCGAATTAATGCTTAATCAAGAAATGATCTGTTT 581
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 302 ATCTTAAATTAATTTTCTTTTCAATTAATTTTGGTATTAATTAATTAATTAATA 361
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 582 ATATACGTAATTTCTATGACATTAAT 609
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 362 ATTCATATCATTTTATTAATTAATTAAT 389
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :

```

## RESULT 21

CNS04D0K/c 945 bp DNA linear GSS 01-SEP-2000

LOCUS CNS04D0K  
DEFINITION Tetradon nigroviridis genome survey sequence 17 end of clone  
101H21 of library G from Tetradon nigroviridis, genomic survey  
sequence.

ACCESSION AL285149  
VERSION AL285149.1 GI:8023560

	KEYWORDS	GSS : genome survey sequence.
	SOURCE	Tetraodon nigroviridis
	ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.
	REFERENCE	1 Roest Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C., Bernic,A., Fitzmes,C., Wincker,P., Brottier,F., Quetier,F., Saurin,W. and Weissbach,J. Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence Nat. Genet. 25 (2), 235-238 (2000)
	JOURNAL MEDLINE	20296633
	PUBMED	10835645
	REFERENCE	2
	AUTHORS	Roeist Crolius,H., Jallion,O., Dasilva,C., Ozouf-Costaz,C., Fitzmes,C., Fischer,C., Bouneau,L., Billaault,A., Quetier,F., Saurin,W., Bernic,A. and Weissbach,J. Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis Genome Res. 10 (7), 939-949 (2000)
	JOURNAL MEDLINE	20359837
	PUBMED	10899143
	REFERENCE	3 (bases 1 to 945) Genoscope. Direct Submission Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequefgenoscope.cnm.fr - Web : www.genoscope.cnm.fr) This sequence is a single read and was generated as part of a large scale clone-and sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cnm.fr/Tetraodon.
	COMMENT	location/qualifiers
	FEATURES	1..945 /organism="Tetraodon nigroviridis" /mol_type="genomic DNA" /db_xref="taxon:59883" /clone="101H21" /clone_1lb="G" /note="Genoscope sequence ID : CORGI01CD11LP1-end : T"
	BASE COUNT	386 a     112 c     96 g     231 t     120 others
	ORIGIN	
	Query Match	8.8% ; Score 54.8 ; DB 29 ; Length 945 ;
	Best Local Similarity	37.2% ; Pred. No.1.2 ; 237 ; Indels 1 ; Gaps 1 ;
	Matches 172 ; Conservative 52 ; Mismatches 237 ;	
OY	159 TCATTTGAACAATTCTACTAATTTTACAGAAATCGTGAGTTACAGC	218
Dd	902 TTATTTTAATAATTTTATTTTAATTTTWTTTTTTTTAMATTTTATTTAN	843
OY	219 ACAGTCAGTAGCTTATTATTATTTATGTGCTCAAGAGATAGCTATTTATGGATCC	278
Dd	842 WTAMWTAATANTTTTATWNTAAATAMAATATWTTTAT-TATTAAAAATTTT	784
OY	279 AATTTACCTTTTANTGCCTGTTAGGCGATGAGATTCATAGTGCAGAAAGTGGCAT	338
Dd	783 TTTATTTAAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT	724
OY	339 ATGCACTAACAAAGCTTTTAAATTTAAATGAGTAATTTTAAATTAATAAAGGTTGAA	398
Dd	723 TTATMAWMTTWTWTWTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTMWM	664
OY	339 AGCAATGTTGCTTCTCAACAACCTTGTCGACTTACGATGATTTAGTTCAAGTAAT	458
Dd	663 WTWMTMMWWMTATATWTTWTATTAATATATTTATTTTATTTATTTATTTTAAAAT	604
OY	459 TGTAATAGTATGTCATATATGATTAATAAAGGAGAGTTTCAAATTTTACAT	518
Dd	603 ATTMTTTTTTAAAAATTTTATATTTATTTTATTTTAAAAATATAAATAATWTTAWMA	544

Oy		519	ATATTATGCAAGTCTCTCAGGTGGCGAGATTAAATGCTTAATACAGATAAGACTTG	578
Dd		543	ATTTTATTATATTTTTTTTTAAAAAAMMTWTATTTTATTATTTTWTTTTMAAATWNTT	484
Oy		579	TTTATATACGTAATTCCTATGCACAATTATTCAGTCGACGA	620
Dd		483	AATTTTATTTATATATAAAAAAAAAAAMAAAAMMTDAAMRNTANTTAA	442
RESULT 22				
AL536104				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				
FEATURES				
source				
BASE COUNT				
ORIGIN				
Query Match				
Best Local Similarity				
Matches 126;				
Oy		159	TCATTTGAAACAATTTTCTACTATATATTTTACAGATCATTTCCGTTGAGTTGACG	218
Dd		776	TTATRRMATAATTTTMTTWTTTWTATTAATTAATWATATATWTAAMTTTTTATATAT	835
Oy		219	ACAGCTAGATATGCTTATTTATTTATATGTTGCTCAGAAGATAGCTATTTTATGATGC	278
Dd		836	AMAAAMMTTWAATATWTTTATTTTAAATAATTTATTTTATTTTAAATTTTAAATTT	895
Oy		279	AATRTTACTTTTATGCTGTACGGATGAGATTTCTATTCATCGGTACAACTGGTGAT	338
Dd		896	TTTATTTATATAMTTTWTAMAAMWATTTWWMAATWATTAGCTAMWMAWMAATATAT	955
Oy		339	ATGACATACAAAGGTTTTTATTTAAATGAGATAATTAATATATATAAAGGCTTGA	398
Dd		956	AMGRRTAATAAAMMTTAAAMATATATATATAAAMAAATWMAAATATATTTATAT	1015
Oy		399	AGGAAGTTTCTTCTCAACAAACACTTGTTGACCTTTTACGTAGTTTAGTCAAGTAAT	458



Dd		1016	TTTWTMTTWTAATSTATAWAAAAAAMTMTTTTTTTAACTMTAWATTMTWTAAMWWAN	1075
Oy		459	TGTTAAGATGTTCGCATCATACATATATAAGAAGACTTTCAACTTTTACCATT	518
Dd		1076	WRITMTWTTARAAAAMMAMAIAFTTTTRIDMAAAAAAAAAAAAAADDAATMMWTTGGAT	1135
OY		519	ATTATATGAACAATCCCTTGCGTGGCAGATATATAGCTTATATACAGAAATAGTACTGT	578
Dd		1136	ATTATMTAMATWTTWTAMDMDATTTTTWMATNTATNMWMTNGTKTNTATNATATAAARA	1195
OY		579	TTTATA 584	
Dd		1196	WA DRMW 1201	
RESULT_23	CNSNOEVL/c			
LOCUS	CNSO0EVL	1101 bp	DNA	linear GSS 04-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC:			
	BACR29B23 of RPci-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.			
VERSION	AJ069706			
KEYWORDS	AJ069706.1 GI:4949849			
SOURCE ORGANISM	GSS. Drosophila melanogaster (fruit fly) Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Epiphytoidae; Drosophilidae; Drosophila. 1 (bases 1 to 1101)			
REFERENCE	Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequençage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  - Web : www.genoscope.cns.fr )  Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osagawa and Aaron Mammossier at Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPcI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain YZ; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPac Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.			
COMMENT	Location/Qualifiers 1..1101 /organism="Drosophila melanogaster" /mol_type="genomic DNA" /db_xref="taxon:7227" /clone="BACR29B23" /clone_lib="RPcI-98" /note="end : T7"			
FEATURES	source			
BASE COUNT	419 a                91 c                60 g               239 t             232 others			
ORIGIN				
Query Match		8.8%	Score 54.6;	DB 29; Length 1101;
Best local Similarity		33.9%	Pred. No. 1.3;	Mismatches 182; Indels 1; Gaps 1;
Matches 131; Conservative		73;		
OY		223	TCAGATATAGCTTATTTATTTTAGTTCCTCACAGATAGCTATTTTATGATCCAATA	282
Dd		1101	TTWWWATWTTWTTWTTTTTTMTWTATMTATATATATTAATTTWTTWTAATATMTAA	1042
OY		283	TTACTTTTATTCGCTTGACGCCATGATTTTCATTAGCTGTACAAAGTTGGTCATATG	342
Dd		1041	TWATATATATTTTWTAAATATATMTWTAAATATATATATATATATATATATATATAT	982

[illegible]



DB	LOCUS	DEFINITION	FEATURES
684	MMMMATWMTATTTTTT	YUHTTHNNHYTWNNHNNYCSCHMTTWTMNCSTTHYCC	743
OY	287	TTTTATTCGCTGTACGCGATTCATTACGCTACAGTTGGTGCATTAGSACTA	346
Db	744	YCMHHWTTTTTYYLNNHTTWTNNHNNHNTTWTTHTHNNMMNNHHTTWTWTYWM	803
OY	347	CAAGGTTTTTAAATTTAAATGAGTAATTTAATATATAAAGGTTGAAGGAAGT	406
Db	804	TTTTTWTTTTTHHTHTHNTHTMTACTHMTATNNMAAAAAMAMTNNHTHTTWTAT	863
OY	407	TTCGCTTCAACAAACACTGTGTACGCTTTTAAAGTAGTTAGTTCAAGTAATGTATAG	466
Db	864	MMYMMATWNNHMMATTTTWTNNHTTMMWTAHTNNHNNCHNTHTHTMMMAAAAAM	923
OY	467	GTAATGTTCCATATATGATTATATAAGAGAGAGTTTCAACTTTTACCATTATATAG	526
Db	924	HTMMMMTWTWTAWNNHMMWTTTNNHTNNHTTCTWNNHYHTTMMHTTWTNNHMYC	983
OY	527	CAAGTCT 534	
Db	984	HHHHMMW 991	
RESULT 25	CNS020K7/c		
LOCUS	CNS020K7	1092 bp DNA linear GSS 01-SEP-2000	
DEFINITION	Tetradodon nigroviridis genome survey sequence T7 end of clone 222L11 of library G from Tetradodon nigroviridis, genomic survey sequence.		
ACCESSION	AL175696.1	GI:7813753	
VERSION	GSS;	genome survey sequence.	
KEYWORDS	Tetradodon nigroviridis		
SOURCE	Tetradodon nigroviridis		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorphi; Acanthopterygii; Percomorphi; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetradodon.		
REFERENCE	1 Roest Crolius,H., Jalllon,O., Desliya,C., Bouneau,L., Fisher,C., Bernot,A., Fitzmes,C., Winkler,P., Brotlier,P., Quetier,F., Saurin,W. and Weisenbach,J. Estimate of human gene number provided by genome-wide analysis using Tetradodon nigroviridis DNA sequence Nat. Genet. 25 (2), 235-238 (2000)		
TITLE	2 10835645		
JOURNAL	Roest Crolius,H., Jalllon,O., Desliya,C., Ozouf-Costaz,C., Fitzmes,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weisenbach,J. Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradodon nigroviridis Genome Res. 10 (7), 939-949 (2000)		
MEDLINE	20359837		
PUBMED	10899143		
REFERENCE	3 (bases 1 to 1092)		
AUTHORS	Genoscope.		
TITLE	Direct Submission		
JOURNAL	Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) This sequence is a single read and was generated as part of a large scale clone-and-sequencing project of the Tetradodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetradodon.		
FEATURES	location/Qualifiers		
Source	1..1092		
	/organism="Tetradodon nigroviridis"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:99883"		
	/clone="222L11"		
	/clone_lib="g"		

[illegible]

BASE COUNT 342 a /note="end : T7"  
ORIGIN 1 c 104 g 462 t 112 others

Query Match 8.7%; Score 54.2; DB 29; Length 1021;  
Best Local Similarity 38.2%; Pred. No. 1.6;  
Matches 105; Conservative 44; Mismatches 126; Indels 0; Gaps 0;

QY 172 ATTTTCTACTATATTTTACAGATCATTTGCTTGGATTACAGACAGTCATATAT 231

DB 734 AKKGGTGTAAAT 793

QY 232 GCTTAT 291

DB 794 TTTTAT 853

QY 292 ATTGCTGTACGCGATGATTTTATTTACGTTGATGATGATGATGATGATGATGAT 351

DB 854 TTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 913

QY 352 GTTTTAT 411

DB 914 TAT 973

QY 412 TCTCAACAACCTGCTGACCTTTTACGATGTT 446

DB 974 TAAAAAATAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1008

RESULT 27  
AL522840 854 bp mRNA linear EST 22-MAY-2003  
LOCUS AL522840 Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED Homo sapiens

DEFINITION cDNA clone CS0DB009YH19 5-PRIME, mRNA sequence.  
ACCESSION AL522840  
VERSION AL522840.2 GI:31041104  
KEYWORDS EST.

ORGANISM Homo sapiens (human)  
SOURCE Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Li W.B., Gruber C., Jessee J. and Polayes D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished

COMMENT On Feb 13, 2001 this sequence version replaced gi:12766333.

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 9464.r For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0DB009CD10Q1&cluster=9464.r. Contact :

Feng Liang Email : fliang@life.techn.com URL :

http://fulllength.invitrogen.com/Invitrogen Corporation 1600

Paradey Avenue Genoscope sequence ID : CS0DB009CD10Q1.

FEATURES  
SOURCE  
1.854  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DB009YH19"  
/tissue\_type="NEUROBLASTOMA COT 10-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 342 a /note="end : T7"  
ORIGIN 1 c 104 g 462 t 112 others

Query Match 8.6%; Score 53.8; DB 9; Length 854;  
Best Local Similarity 35.2%; Pred. No. 1.9;  
Matches 148; Conservative 50; Mismatches 223; Indels 0; Gaps 0;

QY 189 TACAGATCATTTCCGTTGAGTTACAGACAGTCAGTATATGCTTATATATATAT 248

DB 361 TCCAAAMCCCCBBGTTKKKKKKKKTTACCTTTTTTTTTTTTTTTTTTTTTTTT 420

QY 249 TCGTCAAGATAGCTATTTTATTTGATGCCAATATTTTATTTTATTTGCTGTTACGCGATG 308

DB 421 TGGKTATAGGTTNNNGGKTTTANNNTANCCNNCCNNNNNTTTTATATATATAT 480

QY 309 GATTTTCATAGCGTACAGTGGTGCATTTATGACCTCAAAAGTTTTTATATATAT 368

DB 481 AATAATATATTTTAT 540

QY 369 GAGTAAATTTAT 428

DB 541 TTAATATATTTTAT 600

QY 429 TCGACTTTAGCTAGTTTATGTCAGATATTTGTTATGTTTCCATATATATAT 488

DB 601 WATATTTTAAAT 660

QY 489 AAAAGAGAGTTTCAACTTTTACATTTATATATATATATATATATATATATATAT 548

DB 661 AATAAT 720

QY 549 TTAATGCTTAAATACAGATATGTTGTTTATATATATATATATATATATATATAT 608

DB 721 AAT 780

QY 609 T 609

DB 781 T 781

RESULT 28  
BX338020 1076 bp mRNA linear EST 02-MAY-2003  
LOCUS BX338020 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA

DEFINITION clone CS0D1054YA04 5-PRIME, mRNA sequence.  
ACCESSION BX338020  
VERSION BX338020.1 GI:30339670  
KEYWORDS EST.

ORGANISM Homo sapiens (human)  
SOURCE Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Li W.B., Gruber C., Jessee J. and Polayes D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 3701.f For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0D1054BA02Q1&cluster=3701.f. Contact :

Feng Liang Email : fliang@life.techn.com URL :

http://fulllength.invitrogen.com/Invitrogen Corporation 1600

Paradey Avenue Genoscope sequence ID : CS0D1054BA02Q1.

FEATURES  
SOURCE  
1.1076  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1054YA04"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_id="Homo sapiens PLACENTA COT 25-NORMALIZED"











http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CSODI058B05NPL&cluster=3221.r. Contact :  
 Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Paradey Avenue Genoscope sequence ID : CSODI058B05NPL.  
 Location/Qualifiers

# FEATURES

source

1. 999

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CSODI058B05NPL"

/issue\_type="PLACENTA COT 25-NORMALIZED"

/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-cligo (dr)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 331 a 73 c 78 g 216 t 301 others

## ORIGIN

### Query Match

Best Local Similarity 26.7%; Pred. No. 3.2; Length 999;  
 Matches 105; Conservative 103; Mismatches 185; Indels 0; Gaps 0;

```

Oy 131 TTGCTATTATTCAGCGCTAATGGAGCGTCAATTTGAAACAATTTCTACTATATTTTGA 190
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 489 TBGTTNTTTTANBNBEBGGAABDWTATTAKADKKADAAAANGRAKATTTATTTTGA 548
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Oy 191 CAGATCATTCGCTTACGTTACAGACAGTCGTAATGCTTATTTATTTATTTATGTTG 250
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 549 WTTTATTTTNTTTTWTWMMWMAWMAWMAWTTTWTWTTAKKDKKAKAKAA 608
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Oy 251 CTCAGATGATGCTATTTATTTATGATGCTATTTATTTATTTATTTATTTATTTATTTG 310
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 609 DADAAAADADAAAADAAAADAAAADAAAADAAAADAAAADAAAADAAAADAAAADAAAAD 668
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Oy 311 TTTCATTACGTTACGTTACGTTACGTTACGTTACGTTACGTTACGTTACGTTACGTTAC 370
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 669 NKAADAKKKKAAAAAARRAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAA 728
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Oy 371 GTAAATTTAATTAATAAAGGTTGAAAGGATTTGCTTCTCAACAACACCTGTTTC 430
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 729 KKAARAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAA 788
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Oy 431 GACTTTTACGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTT 490
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 789 TAAATTTTAAATTAATAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 848
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Oy 491 AAGAGAGCTTTTCAACTTTTTCACATTATAT 523
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 849 ANAAAAATWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAW 881
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||

```

## RESULT 37

CNS016CO/c

LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC

DEFINITION BACN15E04 of DrosBAC library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

ACCESSION AL106578.1 GI:5622626

VERSION AL106578

KEYWORDS GSS.

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster

REFERENCE Eukaryota, Metazoa; Arthropoda; Hexapoda; Insecta; Pezomyza;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1200)

Genoscope.

Direct Submission

TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :

JOURNAL BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

## COMMENT

Determination of this BAC-end sequence was carried out as part of a  
 collaboration with the European Drosophila Genome Project (EDGP) -  
 http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC  
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
 d'Etude du Polymorphisme Humain) with funding provided by a MRC  
 project grant. The DNA was prepared from embryos by Alain Boucheton  
 and Genevieve Payan. It has been constructed in the vector  
 pBelobAC11.

# FEATURES

source

Location/Qualifiers

1. 1200

/organism="Drosophila melanogaster"

/mol\_type="genomic DNA"

/db\_xref="taxon:7227"

/clone="BACN15E04"

/clone\_lib="DrosBAC"

/plasmid="pBelobAC11"

/note="end : T7"

BASE COUNT 351 a 188 c 167 g 307 t 187 others

## ORIGIN

### Query Match

Best Local Similarity 33.3%; Pred. No. 3.2; Length 1200;  
 Matches 115; Conservative 64; Mismatches 166; Indels 0; Gaps 0;

```

Oy 162 TTTGAAACAATTTCTACTATTTTTCAGAAATCAATTCGTTTGAGTTACAGACA 221
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 1174 WMTAATWMAWMTATTAATATTAATWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAW 1115
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Oy 222 GTCAGTATTCCTTATTTATTTATTTGTTCTCAAGATACGTTATTTATGATCCAT 281
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 1114 TWMTWTWMTWMTATTAAMWTTTWTWTTTWTWTTTWTWTTTWTWTTTWTWTTTWTW 1055
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Oy 282 ATTACTTTTATTCGTTACGAGATGATTCGTTACGTTACGTTACGTTACGTTACG 341
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 1054 TTMWMTWMTWMTWMAATWTTTATTTATTAATWTTATWTTATWTTATWTTATWTTA 995
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Oy 342 GACTACAAAGCTTTTAAATTTAATGAGTAATTTAATTAATAAAGGTTGAAAG 401
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 994 AAAAAAAMTTTTTWTAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATA 935
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Oy 402 AATGTTGCTTCTCAACAACATTTGCGATTTACGTTACGTTACGTTACGTTACG 461
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 934 WMAATTAATTTTAAATTTTTCGTTAAATTTATTCGTTATTTTATTTTATTTTATTT 875
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Oy 462 TATAGTATTTGCTCATATATGATTAAAGAGAGTTTCA 506
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 874 ATWTTTWTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTW 830
    ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||

```

## RESULT 38

CNS06OXV/c

LOCUS T7 end of clone AW0A009H09 of library AW0A from strain CLIB 89 of

DEFINITION Yarrowia lipolytica, genomic survey sequence.

ACCESSION AL411257

VERSION AL411257.1 GI:12180512

KEYWORDS GSS.

SOURCE Yarrowia lipolytica

ORGANISM Yarrowia lipolytica

REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetaceae; Dipodascaceae; Yarrowia.

1 (bases 1 to 759)

Genoscope.

Direct Submission

TITLE Submitted (05-JUL-2001)

JOURNAL JOURNAL

MEDLINE 20584711

PUBMED 11152876

Yeast species for molecular evolution studies

1. A set of

Yeast species for molecular evolution studies

Yeast species for molecular evolution studies

Yeast species for molecular evolution studies



[illegible]

	BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL069440
VERSION	AL069440.1 GI:4949583
KEYWORDS	GSS.
SOURCE	Drosophila melanogaster (fruit fly)
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1101)
REFERENCE	Genoscope.
AUTHORS	Direct Submission
TITLE	Submitted (02-JUN-1999) 'genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE [E-mail : seqret@genoscope.cns.fr
JOURNAL	
COMMENT	- Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oosawa and Aaron Mammoler in Pieter de Jong's Laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y <sup>2</sup> ; cn bw sp, the same strain used for the library p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library or filters for hybridization from the BACPIC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES	location/Qualifiers
Source	1..1101 /organism="Drosophila melanogaster" /mol_type="genomic DNA" /db_xref="taxon:7227" /clone="BACR29P01" /clone_1lb="RPCI-98" /note="Tend : TET3"
BASE COUNT	366 a 66 c 104 g 351 t 214 others
ORIGIN	
Query Match	8.4%; Score 52.4; DB 29; Length 1101;
Best Local Similarity	35.5%; Pred. NO. 3.5;
Matches	151; Conservative 65; Mismatches 203; Indels 6; Gaps 1
Oy	167 AAACAATTTTCACATATTTTTCAGAGATCATTCGGTTGAGGTACAGCAGCTAG 226
Db	552 AAMWMAATTTTTTAAMWATAAATAAAATAATTATTTTAAATTATATATTAAMGMW 611
Oy	227 TATATGCTTTATTTATATGTTTGCTCAAGATAGACTATTTATGTAGCCATATTC 286
Db	612 AAAAAAAMWTAAAHMTAAATTAATTATATMTWAAMATTTAAATAATATTTAAATTA 671
Oy	287 TTTTATTTGCTGTACGCAGTAGATTTCAATACGTCACAGTGGTGATTAAGACGA 346
Db	612 WTWTTTAAAMWTTTAAATTAATAAAMWAAATTTAAAAAATTTWAATAAATAATAAA 731
Oy	347 CAAGAGTTTAAATTTAAATGAGTAATTTAATATATAAAGATGTAAGAAGATGT 406
Db	732 ATAAATTTTTTTTATTTAAAMWTTTAAATTTATWATAAAMWTTWMAAAAAATATW 791
Oy	407 TTGCTTTCACAA-----ACACTGTTCGACTTTACGATGTTAGTTCAGATATG 460
Db	792 TATAAATTTWWGAATAATTAATAAMWTTTAAATATATPAATTAATTTAAATTTTWT 851
Oy	461 TTATAGATTTGTTCCATATATGATTTATAAAGAGAGTTTCAAACCTTTAACATAT 520
Db	852 AATTTAAAMTTTTTTTAAWTAAMATGTAATAAAATTAATTTAAATAAAMATGTTW 911
Oy	521 ATTATGCCAGTCCCTCAGGTGGCGAGATTATATATGCTTATACAGGAATAGTATGTT 580
Db	912 ATTWMTGTATWTTTATTTTAAATGTATATATAAAAAAMWATTAATGATATAMWAAATRTW 971

QY	581	TATAT	585
Db	972	TGTAT	976
RESULT 40			
CNS00DKY	928 bp	DNA	linear GSS 04-JUN-1999
LOCUS			
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC #		
BARC27A24 of RPCT-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.			
AL071865			
AL071865	GI:4948170		
SSS.			
Drosophila melanogaster (fruit fly)			
Drosophila melanogaster			
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
1 (bases 1 to 928)			
Genoscope.			
Direct Submission			
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :			
Bp 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr			
- Web : www.genoscope.cns.fr)			
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).			
The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazuo Oosawa and Aaron Mammose in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCT-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .			
Location/Qualifiers			
1. 928			
/organism="Drosophila melanogaster"			
/mol_type="genomic DNA"			
/db_xref="taxon:7227"			
/clone="BARC27A24"			
/clone_lib="RPCT-98"			
/note="end : T7"			
BASE COUNT	262 a	70 c	84 g 321 t 191 others
ORIGIN			
Query Match	8.4%	Score 52;	DB 29; Length 928;
Best Local Similarity	30.1%	Pred. No. 4.2;	Mismatches 181; Indels 1; Gaps 1;
Matches 118; Conservative 92; Mismatches 181; Indels 1; Gaps 1;			
219	ACAGTCAGTATATCTTATTTATTTAGTGTCAAGAGATGATTTTATGATGCC	278	
490	AATATTAATATGATATTAATTTTATKAWMAAAMAAATTTTWTWTTTTMAA	549	
279	AATATTTCTTTATTTGCTGTAGCGCATGATTTCAATGATGACAGTTGGCAT	338	
550	AMTATTTTWTWTWTWTWTWTWTWTWTATTTAAATTTATTTAAATTTTATA	609	
339	ATGATCAACAAGGTTT-TAAATTTAAATGAGTAATTTAATTTAATAAGGTTGA	397	
610	WTAAATTAATTAATTAATTAATTAATTTATTTTATTTATTTATTAATAA	669	
398	AAGAAATGTTTGTCTTCAACAAACATTTGTGACCTTTACGTAGTTAGTTCA	457	
670	AAAAAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT	729	
458	TTGTTATAGATTTGTCATATATGATTAATAAAGAGAGTTTCAACTTTTACAT	517	

Db 730 WATAMAAAATATWMMWWMAATWTWDGNNNNNMMWWWMMWWMAAAMMWWAAAWMMWWMMWW 789

Oy 518 TATATTATGCAGTCTTCAGGTGGCAGATTAATCTTAACAGCAATAGTACTTG 577

Db 790 AAAAAAAAAAAAAAAMDDDDDDDDMMWAAAKKKKKKKKKKKKKGKKTK 849

Oy 578 TTTTATATACGTAATTCCTATGACAATTAAT 609

Db 850 TKTTTGARMTTTTTTTTTTTTTTTTTTTT 881

Search completed: October 4, 2003, 05:13:17  
Job time : 1941 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 3, 2003, 18:26:46 ; Search time 2052 Seconds

(without alignments)  
12400.481 Million cell updates/sec

Title: US-10-010-160-1

Perfect score: 622

Sequence: 1 atgctgatgatccagctaa.....aattatcgacgctgcagatc 622

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hcg.\*  
3: gb\_in.\*  
4: gb\_on.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_scs.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_scs.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_hcg\_hum.\*  
31: em\_hcg\_inv.\*  
32: em\_hcg\_other.\*  
33: em\_hcg\_mus.\*  
34: em\_hcg\_pln.\*  
35: em\_hcg\_rod.\*  
36: em\_hcg\_mam.\*  
37: em\_hcg\_vrt.\*  
38: em\_sy.\*  
39: em\_hgtgo\_hum.\*  
40: em\_hgtgo\_mus.\*  
41: em\_hgtgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	80.8	13.0	300150	1 AP004598	AP004598 Oceanobac
2	69.6	11.2	299511	1 AE015941	AE015941 Clostridi
3	68.8	11.1	16343	3 AMFGENOM	LM6178 Apis mellif
4	62.4	10.0	303249	1 AP001515	AP001515 Bacillus
5	61.6	9.9	1105	1 BSFLHB	X74121 B.subtilis
6	61.6	9.9	208780	1 BSUBB009	Z99112 Bacillus su
7	57.8	9.3	110000	2 PPMALBPI_01	Continuation (2 of
8	56.2	9.0	84563	5 AC096885	AC096885 Danto rer
9	54.8	8.8	8056	6 AX598900	AX598900 Sequence
10	54.8	8.8	9052	6 AX345047	AX345047 Sequence
11	54.6	8.8	19237	3 AC116967	AC116967 Dictyoste
12	54.4	8.7	8056	6 AX599046	AX599046 Sequence
13	54.4	8.7	347050	3 PPA929351	AL929351 Plasmodu
14	54.2	8.7	310003	1 AE014016	AE014016 Buchnera
15	54.2	8.7	347050	3 PPA929351	AL929351 Plasmodu
16	53.8	8.6	257757	3 AE014837	AE014837 Plasmodu
17	53.6	8.6	1501	6 AX598988	AX598988 Sequence
18	53	8.5	945	6 BD092427	BD092427 Identific
19	53	8.5	10525	1 AE001502	AE001502 Helicobac
20	53	8.5	10860	1 AE000589	AE000589 Helicobac
21	52.8	8.5	7167	6 AX345303	AX345303 Sequence
22	52.8	8.5	11258	1 AE010546	AE010546 Fusobacte
23	52.8	8.5	106763	9 AP002091	AP002091 Homo sapi
24	52.6	8.5	254050	3 PPA929358	AL929358 Plasmodu
25	52.4	8.4	99003	2 AL390756	AL390756 Homo sapi
26	52.2	8.4	133112	2 BX323887	BX323887 Danto rer
27	52.2	8.4	341050	3 PPA929357	AL929357 Plasmodu
28	52.2	8.4	349980	6 AX344554	AX344554 Sequence
29	52	8.4	5297	6 AX251268	AX251268 Sequence
30	52	8.4	40324	6 AX458634	AX458634 Sequence
31	52	8.4	40324	6 AX458634	AX458634 Sequence
32	52	8.4	217879	2 EX005264	EX005264 Danto rer
33	51.8	8.3	5523	6 AX345066	AX345066 Sequence
34	51.8	8.3	15575	3 AY242996	AY242996 Antherea
35	51.8	8.3	73334	6 AX347026	AX347026 Sequence
36	51.8	8.3	73334	6 AX356493	AX356493 Sequence
37	51.8	8.3	110000	2 PPMALBPI_12	Continuation (13 o
38	51.6	8.3	63513	2 AC074061	AC074061 Homo sapi
39	51.6	8.3	119555	9 EX004827	EX004827 Human DNA
40	51.4	8.3	10807	1 CPU88070	U88070 Chlamydophi
41	51.4	8.3	16351	1 AE001283	AE001283 Chlamydia
42	51.4	8.3	300512	1 AE016995	AE016995 Chlamydo
43	51.2	8.2	12025	6 AX346201	AX346201 Sequence
44	51.2	8.2	98299	2 AC133155_3	Continuation (4 of
45	51.2	8.2	156551	2 BX284638	EX284638 Danto rer

# ALIGNMENTS

RESULT 1  
AP004598 300150 bp DNA linear BCT 17-MAY-2003  
LOCUS Oceanobacillus iheyensis HTE831 genomic DNA, section 6/13.  
DEFINITION AP004598 BA000028  
ACCESSION AP004598.1 GI:22777144  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM Oceanobacillus iheyensis HTE831  
Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.  
REFERENCE 1  
AUTHORS Lu,J., Nogi,Y. and Takami,H.  
TITLE Oceanobacillus iheyensis gen. nov., sp. nov., a deep-sea extremely  
halotolerant and alkaliophilic species isolated from a depth of 1050  
m on the Iheya Ridge

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE

FEMS Microbiol. Lett. 205 (2), 291-297 (2001)  
21623015  
11750818

Takami, H., Takaki, Y. and Uchiyama, I.  
Genome sequence of *Oceanobacillus iheyensis* isolated from the Iheya  
Ridge and its unexpected adaptive capabilities to extreme  
environments

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE

Nucleic Acids Res. 30 (18), 3927-3935 (2002)  
22220767  
12235376

(bases 1 to 300150)  
Takami, H., Takaki, Y. and Chee, G.  
Direct Submission  
Submitted (26-DEC-2001) Hideto Takami, Japan Marine Science and  
Technology Center, Deep-sea Microorganisms Research Group, 2-15  
Natsushima-cho, Yokosuka, Kanagawa 237-0061, Japan  
(E-mail: takami@jamstec.go.jp,  
URL: http://www.jamstec.go.jp/jamstec-e/bio/DEEPSTA/FResearch.html,  
Tel: 81-468-67-9643, Fax: 81-468-67-9645)

# FEATURES

## source

Location/Qualifiers

1..300150

/organism="Oceanobacillus iheyensis HTE831"

/mol\_type="genomic DNA"

/strain="HTE831"

/db\_xref="taxon:221109"

/note="halocolerant and alkaliphilic"

28..1017

/gene="OB1462"

28..1017

/note="CDS\_ID OB1462"

/codon\_start=1

/transl\_table=11

/product="hypothetical conserved protein"

/protein\_id="BAC13418.1"

/db\_xref="GI:22777145"

/translation="MGVMEHREHYLKEESIKGLAIKPDGIYVDCITGGGSHLTIAS  
RLNNGQLPAPFODDQALAAADRLSTADRIYVQSNFGLAEBOIKKGIHQVIGIL  
FDLGVSSFDLDGDRGFSYNDALIDMRDQSOHLAYEVNEMSEYERLVSIFSYGE  
EKFSQIARKEIAYREODIKITQDLVEIILDAIPAPARKGHPAKRIFOALRIYVN  
DELVEFNNAKHQAARLAVNGRIYAVITFHSLEDRICQAFKKMSTDKPTPLPIYVE  
SHQAFKLVTRKPIYADTSELDENRRSRAKRLVIEKSEMHETFEYEGMKK"

## gene

## CDS

1017..1391

/gene="OB1463"

1017..1391

/note="CDS\_ID OB1463"

/codon\_start=1

/transl\_table=11

/product="hypothetical conserved protein"

/protein\_id="BAC13419.1"

/db\_xref="GI:22777146"

/translation="MSVNHATKEMNTQYRPEYQVAVKVKRTKRWITKGEKVIYAF  
AGVLMIAFVYVWSSYSDTLNRELQIQEGSVQOKIANEGLSPEKEELSPERITIN  
IAENGKIQNTQEVQVQVQSN"

## gene

## CDS

1408..3648

/gene="OB1464"

1408..3648

/note="CDS\_ID OB1464"

/codon\_start=1

/transl\_table=11

/product="penicillin-binding protein 2B (cell-division  
septum)"

/protein\_id="BAC13420.1"

/db\_xref="GI:22777147"

/translation="MKKRTTHMSSIIYIIFVALFLITGRFMYIOAGVSGVQIIN  
EMADQRTASTYTMSEKGIYKQKCVIATYRPTKRAIIDEAYSENAEELHVSNA  
ETIAKCLAPLVGEEETIOEPIQNGLDNNRFQVGRGKQISKQVQKIEIDLNTPGI  
NFEERPMRYPLNGMFASHIIGFAREREKTEGVEHITIGVGMEDVWNDILGGKQY  
ISYQDLVNNKLLDPSSEVLIQLEPDGDDIYLTIDOKITQLLEDVMTQADEKYNPTMNA  
IIMDKSGEIVAMSRPSYVDPNPPTDVENWYNDPISTPFEPGTSWKIIFTMAAIEBGV

## gene

## CDS

3732..5663

/gene="spovD"

3732..5663

/gene="spovD"

/note="CDS\_ID OB1465"

/codon\_start=1

/transl\_table=11

/product="stage V sporulation protein D (sporulation  
specific penicillin-binding protein) (spore cortex)"

/protein\_id="BAC13421.1"

/db\_xref="GI:22777148"

/translation="MKRVSTYITTKRIYVFLALIIYVIGRLAYQVPLQDGLAG  
QANLWTRDITFAERGIYLDSEGEVAENYATPVIYVMPQITNPEQTAVQLKILID  
VPQDRVHETITKNTSSVWIREGKIKIDQKQALRELDLSGVYLAQSKRYPPNGDDL  
SHVGFTEGIDNQGMLSESYDRLKNGSLISYSPAKGKLELDENPVDGILN  
LQTTINRTVOTIIEBELDKATIKYNPQALAIYVNPRTGILGMSRPNPEYQDV  
DASIRGNLPIYMSYBPGSPFKITITLAAIEBQVNLBEDEHDDGDIENGGATTHCW  
KKGCHQDSYIEVYQNSCNPGFVNLGRLGTEKIFSYIDAFGRGKTGIDQGGNGI  
LFOENVPPELATTISFGQVSVTPIQVAVAAVAVNGVLYKRYIADWIDPVNGET  
VEKEPOLQNNVISEDTSBEIRVLESVVAQGTORPAYVDGTVGKGTQKQKPG  
GYMNNYIVSFIQFAPADDEIVVYVAVDNEKDTLQFGVVAAPIVGTIIDSIRAME  
VEPTDGLKEKESWPDQKVEVPLIGESTQKLEVYLVDLSVTVNGEDTIIIEQSPK  
GTMLSGSTVRIYLDENE"

5802..7283

/gene="murE"

5802..7283

/gene="murE"

/BC\_number="6.3.2.13"

/note="CDS\_ID OB1466"

/codon\_start=1

/transl\_table=11

/product="UDP-N-acetyl[muramoyl]alanyl-D-glutamy]-2,  
6-diaminopimelate ligase (EC 6.3.2.13)"

/protein\_id="BAC13422.1"

/db\_xref="GI:22777149"

/translation="MEIKKLEICTFYNNNGEVCNCTISLMSRKITSSAVCIIT  
GFTVGHVYDQAVKNGASAIETSKPLKEGVITQVEDNRLAMIAVYIYPIK  
HFPLIGVINGKTVTYLLDKIPEYHQKAGVIGTIOVKIGESTFPIVNTTPNAL  
QKTFVNRKEDKQVIMEVSHALDMGRVYGCVDIYVFTVLSDHLDYHDDIDYLR  
AKSLIFALQNGYDSEKERYAIINDDSSSHLGRSAOHVITYSCKEATIMAKDIE  
LTASGIRFKLHSPIGNITIOSRLMGNNVNMVLAASAAIASKYPLVITQALSTIEG  
VNGRFPPIVRCQNVSVIYDPAHPPDSILENLOTIKDKAKNNVYVVCGGDRKKRP  
LMAVVALNYADHVAFTSDNRTPEQALIDDMTKELDANSGSVYVYVDRKGIYAKAIQ  
SAQKDDIVLNGKHEHYQIIIGTKYVDFDRVARNAIKQGE"

7289..8656

/gene="murF"

7289..8656

/gene="murF"

/BC\_number="6.3.2.15"

/note="CDS\_ID OB1467"

/codon\_start=1

/transl\_table=11

/product="UDP-N-acetyl[muramoyl]alanyl-D-glutamy]-2,  
6-diaminopimelate-D-alanyl-D-alanyl ligase (EC 6.3.2.15)"

/protein\_id="BAC13423.1"

/db\_xref="GI:22777150"

/translation="MLPTVOMLAEIPKHOGLISTDTITQVTRTPSPNSIYIPI  
VGERFDGHNFIPLDAIKGALATLMDKSIPLSEEDNSFPVFEVDITLMOQLANSYL  
RSVNPYVGVGTSKGTTKDOLITSYVTKTGAHGNLNHGIKPLLETISGQDIE  
VYVEMSGQGEIHELKSTADPYPVYVYTNIGESGRIENLGSRIIAKGLLETISGQD  
NGNIFIDDEBLKKNITDMDSDIIGITGKESDINIYVKKHPDPPTTTEVTPKPH  
VNLGHEHNAANAAFIISGALNLSHNLIDLAENKLEVTQHRFLEAVTISRSIEIY  
YNSLGEHMKASIQVXKEGFRONVLAIGLIEGTYSYHHRISIAVTSRISIEIY  
YGEQAFYIYDEITKQHKVAVVHYIKQEDVAVSIKELHNNOTLILFASGGMKEMFI  
KELQO"

```
gene
8709..9689
/gene="miray"
8709..9689
/gene="miray"
/EC_number="2.7.8.13"
/note="CDS_ID 081468"
/codon_start=1
/transl_table=1
/product="phospho-N-acetylmutamoyl-pentapeptide
transferase"
/protein_id="BACJ3424.1"
/db_xref="GI:22777151"
/translation="MDLTGLITTVAFLLITVLLSPIFPIPLRLNFGSIRERGPQS
HOKKTGPTMGMLIIFSIITSLIMASRTDEGINYQVWLIFVLFYGLGFDDFI
KVAKRNIGLTSKOKLFGOIIIALVFYILNOSFTVIYVPGTELOEDGMFYAVIV
IFMMVGSNVLNFDGLAGTAAIAFGAFIIVANGIPDHVAFAVAVGALLG
FIVNAPAKVPMGDGSLALGAIASILKIEIILIIIGSFVETLSVIQVLS
FKTTGKRVFKNSPLHHRYELLGMSERVITFMVLGLFPAVLGYIEVGM"
9689..11038
/gene="murD"
9689..11038
/gene="murD"
/BC_number="6.3.2.9"
/note="CDS_ID 081469"
/codon_start=1
/transl_table=1
/product="UDP-N-acetylmuramoylalanine-D-glutamate ligase"
/protein_id="BACJ3425.1"
/db_xref="GI:22777152"
/translation="MNVLINPYQHVVLVLGAKSGTAANYLONHIOVTINDGMATL
EDATVQKLTGMAELVGSHEISVLDEKDLIVKNPGIRYDVAIVEAQRIPIVISEV
ELVHVLNOPYIGTGSNGKTTTTLITEMDRNVSNGVAKNIGVATEVASELSQD
EKWMEELSSPOLGIDLOESTAVLNLFEALDHYGSFENVYAKNIEFVSOKHXY
LIYVADDNWSAIIKTAETKVPSSRPADGAMDDDLFYDEKILAIIRDVLVG
KHNENILAAATKALNGATNEGIVQUTLTSSGKRLHEPVGIVNGXYINDSATNI
LATKALAAFNKAVNLLAGLDKNGITBELPYLHAKAVVFESLIGKLDAGVAAV
IPVLEKALDVQHAVEVAFAALTDQDTLLSPACASWQYKTFEERGDMFIQALRLX"
11116..12234
/gene="spvE"
11116..12234
/gene="spvE"
11116..12234
/gene="spvE"
Query Match 13.0%; Score 80.8; DB 1; Length 300150;
Best Local Similarity 48.3%; Pred. No. 7.7e-06;
Matches 226; Conservative 0; Mismatches 242; Indels 0; Gaps 0;

19 AAAACAGAAAGCAACCCCGAAAGCAGCTGAGAGAGCTGCTCGAAGGAGTGTCCCT 78
Db 112665 AAGACAGAGAAAGCAACTCCAAAAAGAGGTTAGTGAACGAAAAAGGTCGAGTAGCT 112724
QY 79 AATCAGAGAGGTTACTAAGCACTGACTGACGAGGAGTCTGGGCTTGCTATT 138
Db 112725 AAAAGCCAGAGTGTAAATACGCGCTTTTGTCTATATAGGCTTTTATACCTTTGTTGTA 112784
QY 139 TATTCAGCGGTAATGGAAGTCAATTTTGAAACAAATTTTCTATATTTTACAGAAATCA 198
Db 112785 TTCGAAATATATGCTGCAATATATACCGGTTTTTTCGAAAAATCATTTACAGAGTAC 112844
QY 199 TTTTGGTTGAGGTTACAGCAGTCACTATATGCTTATTTATTTATTTGTTGCTCAAGAG 258
Db 112845 ATTATTTGGGATCTTCACAGAAAAAAGCTTAATTAAGTTTAATAGTCTTCACCTGAT 112904
QY 259 ATAGCTATTTTATTTAGTCCCAATATTTACTTTTATTTGCTGTTAGCGGATGGAATTTCACTA 318
Db 112905 GCAAGGATGATGTTGGCTCCAAATATATATTTGCTATTTATCACTGCAATGCGACAAAC 112964
QY 319 CGTGTACAGTTGCTGATTTATGACTCAAAAGTTTTTAAATTTAAATGAGTAATTT 378
Db 112965 TTAGGCAAGTTAGGTTTATTTAGTACGAGCATTAATTAATTTGATTTAAAAAAATTT 113024
QY 379 AATATAATAAAGGTTGAAGAAATGTTTGTCTTCAACAAACACTTGTTCGACTTTTA 438
Db 113025 GATCGATTCAGGTCAGAAACGATTTTTCATACGTCGATTAAGTGAATTTTCTGAAA 113084
```

```
QY 439 CGTAGTTAGTCAAGTAATTTATAGTATGATTTCCATATGATTT 486
Db 113085 TCATGTTAAAAATTTGTTTGTTCGCAATGTTACATTTATGATTT 113132

RESULT 2
AE015941/c 299511 bp DNA linear BCT 06-FEB-2003
LOCUS Clostridium tetani E88, section 6 of 10 of the complete genome.
DEFINITION AE015941 AE015927
ACCESSION AE015941.1 GI:28203514
VERSION AE015941.1
KEYWORDS
SOURCE
ORGANISM
Clostridium tetani E88
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
1 (bases 1 to 299511)
Bruggemann,H., Bauner,S., Fricke,W.F., Wierer,A., Liesegang,H.,
Decker,I., Herzberg,C., Martinez-Arias,R., Merkl,R., Henne,A. and
Gotschalk,G.
The genome sequence of Clostridium tetani, the causative agent of
tetanus disease
Proc. Natl. Acad. Sci. U.S.A. 100 (3), 1316-1321 (2003)
2457253
PUBMED 12552129
REFERENCE 2 (bases 1 to 299511)
Bruggemann,H., Bauner,S., Fricke,W.F., Wierer,A., Liesegang,H.,
Decker,I., Herzberg,C., Martinez-Arias,R., Merkl,R., Henne,A. and
Gotschalk,G.
Direct Submission
Submitted (31-OCT-2002) Goettingen Genomics Laboratory, Institute
of Microbiology and Genetics, Georg-August University,
Giselastr. 8, Goettingen 37077, Germany
Location/Qualifiers
1..299511
/organism="Clostridium tetani E88"
/mol_type="genomic DNA"
/strain="Massachusetts"
/sub_strain="E88"
/db_xref="taxon:212717"
/complement (317..1060)
/locus_tag="CTC01397"
/complement (317..1060)
/locus_tag="CTC01397"
/codon_start=1
/transl_table=1
/product="membrane spanning protein"
/protein_id="AAO35956.1"
/db_xref="GI:28203515"
/translation="MYKEVWKKYIKGFLLPFLVQILPKSNGLANATFDLFGK
GSLIENSTIFNOLIINDIIPISLELFGALFYSPIIMTITSLNENKISYGEDEKSI
GFILRIYVNLITISATSMVLIFLGFRTILLPKNVILMTITRIIITPEBAIVYNTSV
GESLIRKGIELGKYSGEIILIGVIGMLGVGWSGIVSTNITTPKTNFIAVIFISFIS
NIQVYFPMFWSICKREKAEKTELELY"
complement (1312..2043)
/locus_tag="CTC01398"
/complement (1312..2043)
/locus_tag="CTC01398"
/codon_start=1
/transl_table=1
/product="membrane spanning protein"
/protein_id="AAO35957.1"
/db_xref="GI:28203516"
/translation="MYKEVWKKYIKGKFLPFLVQILPKSNGLANATFDLFGK
IKESITWANSIGENAIANSITLWFLVSSLFYFLVKSIVNRPINKENF
EESIDYFTRFGVNIIMTIIILAVLVSGFIISPYVLSPILIVMDLTKRCGAY
LVYSIDIDALSDGVLDGQVFWRIIFLSFIITFILNLTASTNLNIIITYSIMDLK
VSIKFIYLMFSWYICKREDVLNRKK"
complement (2209..2793)
/locus_tag="CTC01399"
/complement (2209..2793)
/locus_tag="CTC01399"
```

/note="similar to histone macro-H2A.1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved protein"  
 /protein\_id="AA035958.1"  
 /db\_xref="GI:28203517"  
 /translation="MNNNLNKISIIKGDITKESVDIAIVNANSVLLGGGVGGAITHR  
 AGSGEIIKEKEIKINIKTGPIEGSVITGSGGKAKVYIAVGPIMWGGSNGETLLA  
 NCTYNSNLNLAQEKDITKTIAIPNISTGVYGFPODLAVKIVFTKMEINIEKKDKETIKF  
 VCPDDNMVRLYLITFTFMVNFILNTVGFIKGMT"  
 3266..4852  
 /locus\_tag="CTC01400"  
 /locus\_tag="4852  
 /locus\_tag="CTC01400"  
 /locus\_tag="CTC01400"  
 /codon\_start=1  
 /transl\_table=11  
 /product="periplasmic dipeptide transport protein  
 precursor"  
 /protein\_id="AA035959.1"  
 /db\_xref="GI:28203518"  
 /translation="MLKRSKKNKINPNKLAIIVWSIALSGGTSKSEKDTTKDSNK  
 EIVASRPDVAPEKENVHPNANLVSVMKEGSPPTPKYDKITPYGLIDCVCTDPT  
 KLQGVFPFHGEEVNDIVANFDRIYKELKNGKSNPTPTFKYDKITPYGLIDCVCTDPT  
 VKNQFKPIETPLVAVNVRSPSILFEHNDGKQGVTVKPEPAIEPKLISKKNSTCL  
 LEADVNDWGKRSKSKIRIKTIPBQGRYSALKAEGVNWGLDIAIIPNAIEELKLOS  
 KFEISIKSTKINHFSLGNGKPEFNDPKRKEVASIADRNIIVDKPYKGGNPOMLL  
 NHSVPEPKEMKPEPNLKEKASVLEKCKQNTVMIIIPSEFTKQYKEADQIDSV  
 LKEGIANVNIIDPATDPNINKGDIPLQNHQIDGLPMPEBYTMDVKNVNSGNTKN  
 YHFGKDKDKXNDMLMKLIDETLDEKRAAIYDELQISAKDSTIPLFSENNLYVSKKE  
 VTGHNAFIYGTITPELPELQMSK"  
 4853..5803  
 /locus\_tag="CTC01401"  
 /locus\_tag="5803  
 /locus\_tag="CTC01401"  
 /locus\_tag="CTC01401"  
 /codon\_start=1  
 /transl\_table=11  
 /product="transport system permease, nickel or dipeptide"  
 /protein\_id="AA035960.1"  
 /db\_xref="GI:28203519"  
 /translation="MKESLXLYFKKLIIHAPVLIGVSLISFMTGLIAPGDPALQILTK  
 DSISOPKREIYTLREMGNDPILIOYGRMKAKIKCDLGSYTDPOVDELVKIK  
 PITPLISFANTYKIDGILGIIIVAVNNKSHDNGRIILFISIDGPFPAFIILT  
 IFSELKPLINMGVEGKKNLIMSGFLASGTTGIVIMRLIRATVILKEKNYIIAASK  
 GLSSKTYIVHAFNLSTLPTVLLIGVFGIIGGTAIVETIPALPMGSPAINGIPIPR  
 DYPVIOGVIVFTGIIYIIFPLIIDLVLVLPNKIRIGES"  
 5787..6620  
 /locus\_tag="CTC01402"  
 /locus\_tag="5787..6620  
 /locus\_tag="CTC01402"  
 /locus\_tag="CTC01402"  
 /codon\_start=1  
 /transl\_table=11  
 /product="transport system permease, dipeptide or nickel"  
 /protein\_id="AA035961.1"  
 /db\_xref="GI:28203520"  
 /translation="MEVNLMLKHFRFVITLGLSIFAPLISPYPIKVDANLTL  
 LPASEHNLGTDALGRVSRSLIAGRSSIITLVSASIIITLGLIIGLSSYGXGKLK  
 DSFIIDVNIILQSTSPSSNMIAITIGMREPIKSIILSLVITVSMASPSYSEVLISR  
 PEDYIEGRALGADMPFIITHHIITPNTIGFVLFRIGVILSIALSFLKGLGQRP  
 EPTDGMINDAKTYIISHPLMIAPGATIMLSTISNIIIGDFIRDYIRDKQDYNQPI  
 6634..7575  
 /gene="oppD"  
 /locus\_tag="CTC01403"  
 /locus\_tag="6634..7575  
 /locus\_tag="CTC01403"  
 /locus\_tag="CTC01403"  
 /codon\_start=1  
 /transl\_table=11  
 /product="oligosaccharide transport ATP-binding protein oppD  
 protein"  
 /protein\_id="AA035962.1"  
 /db\_xref="GI:28203521"  
 /translation="MNMDDIKMLKVEPPTKGNVRAVDNISLIKINMGKVLIGETG  
 SKSVYLSIIIGLSSNSKILIGELIYVNTMLISLAINDRNRNLRGVADIIIDQNSPTAL

	gene	NPINRKGQICEBGRNNKDCIKHYVNIILKELEHTLTPNSNCAVSAPSEFSLSGCKORALA SMGIYCPDNMIADPEPKGLDAIPRCQYTTELKTITNSNCNMILLITDMLSKULC NIAMWYGNDIIEGKAYDVLENKRPRYTGVLFSOPHKELQPMKSPBLINPPIGCK FSQRCSNRIDPICLSLKRPDYTLLENGSRVCFLPDKSK"
	CDS	7559..8329 /locus_tag="CTC01404" 7559..8329 /locus_tag="CTC01404" /codon_start=1 /transl_table=1 /product="peptide transport ATP-binding protein"
	gene	/db_xref="gi:28203522" /translation="MIKVNULTKSFSRNRFIRKSLVDNVNSHIIKKGITLTGLIGESG CGEKLTKRIILIKLSCDNGEIIYFNKDITNYSKNINLRKMQIIFQHPDTSLNPSK TILLSLEPIILHNLDKNEAIKIKKYLDYNLSDIELNRYPHOISGGQIORIAILR VLLKPKEVIIDEPSTMDIVSOAQIIQLKLIKOEKFNTFYLFISHDLVLPADFDEI AVMKRKIEIGKNYDIYNPNQEYTKLIHTFKRRD"
	CDS	8364..9224 /locus_tag="CTC01405" 8364..9224 /locus_tag="CTC01405" /codon_start=1 /transl_table=1 /product="hypothetical protein"
	gene	/protein_id="AAO35964.1" /db_xref="gi:28203523" /translation="MLNMNKSTFIQHTKENYGLDFEKLWOEGILAKSYDVORANDPY VEKNWDPAFYKNDKPLVDYVAEADKLEIIEEDKIIAEGPOTGFETLPMSSH KKIIAIDSEYMEILEINKKIISKNIIILNKSIDASKEDATIEKDTIFENVALYRMNI KDSLIRNNLAKEKRVYVMTLORSPPNPFICKIQQGGLKTRSPYIIONILVELGIDS NIEFDIITKPIQYNNKEDLYKSFTEYKGSPLPIDNIKLLDQVITYKQYIIFNNKLK VAFIHMTLTK"
	CDS	complement(9289..10209) /locus_tag="CTC01406" complement(9289..10209) /locus_tag="CTC01406" /codon_start=1 /transl_table=1 /product="putative transposase"
	gene	/protein_id="AAO35965.1" /db_xref="gi:28203524" /translation="WKLLTOCKYTLIKSVIIDKYKIKNLVSLCLKSNVSRSGYYNES IKSHNIRSEREDISTNNITLIAYNFRKAKKARQIKMTLENQHIANVRIM HXYDIIICHRKANPYRRMKNKESVLPVNLNREFKODVPGLVLTDTIYLFXAGQ RAYLETVDJGSTEINELLTYNLSKSLKIDIVETIEKLVSNNYIISDSFLHSQGAHY TSYIPOKLLKKYKNIKOSMRPGCMDNAPOSFEFGHLKETINIKCNTFAELLEIDE VWDYANNRGOMNLKQMPYKVRNHLLSTA"
	CDS	complement(10206..10625) /locus_tag="CTC01407" complement(10206..10625) /locus_tag="CTC01407" /codon_start=1 /transl_table=1 /product="hypothetical protein"
	gene	/protein_id="AAO35966.1" /db_xref="gi:28203525" /translation="MSKKIFTENEAILSENKFFVKVYSCSGITTYTBFKRLPTMENKV GEPRIFEECGFINIDIIGLRVOSCGRMAARSGVANQLODRKENTGRPTERKNL PLEEKYEKLOTOKILLEANEILLKLETERLSVAKKK"
	CDS	complement(10779..12323) /locus_tag="CTC01408"
	gene	
Query Match	11.2%; Score 69.6; DB 1; Length 299511;	
Best Local Similarity	48.0%; Pred. No. 0.00084;	
Matches 290; Conservative	0; Mismatches 309; Indels 5; Gaps 3;	
Df	269487	TAAAACGGAAGCCACTTCCTCAATAATTGTGAGGAACTGAAAAAGGACAGTAGC 269422
Df	18	TAAAACGGAAGCAACC CGCAAACGAGCTCAGAAGCTCGTTCTGAAGGAGATGCC 77
Df	78	TAAATCAGAAAGCTTAAGCATTTGACTACTGCACAGAGAGATGCGGCTTGCTAT 137





```

CDS
    /product="rRNA-Asp"
    /note="codon recognized: GAC"
    /anticodon=(pos:4325..4327,aa:Asp)
    4370..4438
    /product="rRNA-OTHER"
    /note="codon recognized: AAU; anticodon unspecified"
    4444..4602
    /codon_start=1
    /transl_table=5
    /product="rRNA-Asp subunit 8"
    /protein_id="AAB96801.1"
    /db_xref="GI:829006"
    /translation="MPQMMPMKMFILYIYLLIFLFLIMLSMLIKIKKELIKIK
    LKKNMFW"
    4584..5264
    /codon_start=1
    /transl_table=5
    /product="rRNA-Asp subunit 6"
    /protein_id="AAB96802.1"
    /db_xref="GI:552442"
    /translation="MKLLMMNLFFMPDPSTSNLSNNLFFMLPILIPFSLFPLIOS
    RIMFLMKTLMFNFNFVSKSKYQSNIIIFISIMLYIMINIFSLIPYFETLSHL
    LLMFLSLTFMPSFLIYLYNNYIMFLSHLVNLPVFLMFMVILIELISILIRPWL
    STRLSANLISGHLITLIGLIFISNISILIPIMLQNNLLEIFMSMIGSVFSLI
    LLYSESNN"
    5285..6064
    /codon_start=1
    /transl_table=5
    /product="cytochrome oxidase subunit 3"
    /protein_id="AAB96803.1"
    /db_xref="GI:552443"
    /translation="MKKNEPFHMTVNSPWPILISGSPNNTLISYIYIYSSISMMIL
    NFINSILMMWFRDILIRESTFGHMSMTITFLKSMFLISLMEFISFPMFFH
    SSISNIEINMTPEPKIKFENPMELIPINSPIVSSGFTVLISHYVILINLKLSKS
    YLLTLILGIFLTIQTIYSSNFCFNDISYGSIFPMATGFHGLHVLIGSIFLLISL
    YMMNHHFSNNMNFELAIWYMHFVDVIMFLYETFLYLLI"
    6119..6184
    /product="rRNA-Gly"
    /note="codon recognized: GGA"
    /anticodon=(pos:6149..6151,aa:Gly)
    6185..6538
    /codon_start=1
    /transl_table=5
    /product="NADH dehydrogenase subunit 3"
    /protein_id="AAB96804.1"
    /db_xref="GI:552444"
    /translation="MKFIFMVFIPIFILLISILLINKFISYKKDYKSPFEGFN
    PITKANLPISLPFLMTMWFILFDEVLILFLPIFYKSSSTMSIYLMISIFLLILIT
    TILLEMNNYLMFLP"
    complement(5572..6638)
    /product="rRNA-Arg"
    /note="codon recognized: AAG"
    /anticodon=(pos:6602..6604,aa:Arg)
    6734..6802
    /product="rRNA-Asn"
    /note="codon recognized: AAC"
    /anticodon=(pos:6765..6767,aa:Asn)
    complement(6810..6878)
    /product="rRNA-OTHER"
    /note="codon recognized: AUU; anticodon unspecified"
    complement(6892..8556)
    /codon_start=1
    /transl_table=5
    /product="NADH dehydrogenase subunit 5"
    /protein_id="AAB96805.1"
    /db_xref="GI:829007"
    /translation="WIKMWCGILLPEFSFLMMLSLYLLYLNKKEFFPMNTYNSM
    KFNFLIIDKSLMFLVLSMIFSMIITYISYWDLSLMDRFLYLLMLPLISMFL
    ILSPNNLSILIGDGLISYCLVIYMKKSPSGAVTILLRLDGLILMGLMT
    YGSMNLSFYKMEPMIYILMAFKAQIPSTWLPNMAAPPVSSIVHSSTLVT
    AGIYLIRVNLDPNKKYIMLIASITMFLAGLVANFELDILKVVAYSLQLGFM
    SMISISTEVLPHLPIHMAKSIEMCVSNVHWNODIDEMYGMYIYPMKSMI
    LIPSILSLCGPFLVGYSKDILIEFFSKMIFPMINILIGITITVTSIPMILVL

```

```

CDS
    /product="rRNA-His"
    /note="codon recognized: CCA"
    /anticodon=(pos:8589..8591,aa:His)
    complement(8644..9987)
    /codon_start=1
    /transl_table=5
    /product="NADH dehydrogenase subunit 4"
    /protein_id="AAB96806.1"
    /db_xref="GI:552446"
    /translation="MYLLLLIMLMNLMMSIYLFMLFMKMKNNLILIGNLIINL
    LNLFNINWIDWYIRCONLSNNYSGLMLTLMIFGLIPISLNNNSLNCFLNNLLM
    ISLLVFLSNLLFLFYFERGLILFYLYVKKGYSENPMLSGFYMLFISLPL
    YIYIYLYIDYSLNFMLEMLNINLNMLIYILMSFLYKIPYILPHGMLIKHVEP
    YGSMILASIMLKLGGMRLMIYKNEBILIOKLIMVINSFGLVLSLMSQFDM
    KSIIAISIVHGLIMISMMTFLKISLIGYLMMLISGSSGLFVLNVIYQTSNR
    LMFINKMINPMSMLLMFLMSNNGSVSNLISEVMLLIGMISMLKFMILLMM
    YCLPSIYISYLFMLPIHNGKIFIMFKIKNGILVEYFLLHWMIPLMMLFKLYFI"
    complement(9991..10254)

Query Match. 11.1%; Score 68.8; DB 3; Length 16343;
Best Local Similarity 47.6%; Pred. No. 0.0027;
Matches 202; Conservative 0; Mismatches 222; Indels 0; Gaps 0;

CDS
    /product="rRNA-His"
    /note="codon recognized: CCA"
    /anticodon=(pos:8589..8591,aa:His)
    complement(8644..9987)
    /codon_start=1
    /transl_table=5
    /product="NADH dehydrogenase subunit 4"
    /protein_id="AAB96806.1"
    /db_xref="GI:552446"
    /translation="MYLLLLIMLMNLMMSIYLFMLFMKMKNNLILIGNLIINL
    LNLFNINWIDWYIRCONLSNNYSGLMLTLMIFGLIPISLNNNSLNCFLNNLLM
    ISLLVFLSNLLFLFYFERGLILFYLYVKKGYSENPMLSGFYMLFISLPL
    YIYIYLYIDYSLNFMLEMLNINLNMLIYILMSFLYKIPYILPHGMLIKHVEP
    YGSMILASIMLKLGGMRLMIYKNEBILIOKLIMVINSFGLVLSLMSQFDM
    KSIIAISIVHGLIMISMMTFLKISLIGYLMMLISGSSGLFVLNVIYQTSNR
    LMFINKMINPMSMLLMFLMSNNGSVSNLISEVMLLIGMISMLKFMILLMM
    YCLPSIYISYLFMLPIHNGKIFIMFKIKNGILVEYFLLHWMIPLMMLFKLYFI"
    complement(9991..10254)

RESULT 4
AP001515/c 303249 bp DNA linear BCT 01-FEB-2001
LOCUS
DEFINITION Bacillus halodurans genomic DNA, section 9/14.
ACCESSION AP001515 BA000004
VERSION AP001515.2 GI:12641879
KEYWORDS
SOURCE
ORGANISM Bacillus halodurans
Bacillus halodurans
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE
1 (sites)
Takami,H., Nakasone,K., Hirama,C., Takaki,Y., Masui,N., Fujii,F.,
Nakamura,Y. and Inoue,A.

```

TITLE An improved physical and genetic map of the genome of alkaliphilic  
JOURNAL Bacillus sp. C-125  
MEDLINE Extremophiles 3 (1), 21-28 (1999)  
PUBMED 99184645  
REFERENCE 10086841  
AUTHORS 2 (sites)  
TAKAMI, H. and HORIKOSHI, K.  
TITLE Reidentification of facultatively alkaliphilic Bacillus sp. C-125  
JOURNAL Biosci. Biotechnol. Biochem. 63, 943-945 (1999)  
REFERENCE 3 (sites)  
AUTHORS TAKAMI, H., NAKASONE, K., OGASAWARA, N., HIRAMA, C., NAKAMURA, Y.,  
MASUI, N., FUJI, F., TAKAKI, Y., INOUE, A. and HORIKOSHI, K.  
TITLE Sequencing of three lambda clones from the genome of alkaliphilic  
JOURNAL Bacillus sp. strain C-125  
MEDLINE Extremophiles 3 (1), 29-34 (1999)  
PUBMED 99184646  
REFERENCE 4 (sites)  
AUTHORS TAKAMI, H., TAKAKI, Y., NAKASONE, K., HIRAMA, C., INOUE, A. and  
HORIKOSHI, K.  
TITLE Sequence analysis of a 32-kb region including the major ribosomal  
JOURNAL protein gene clusters from alkaliphilic Bacillus sp. strain C-125  
MEDLINE Biosci. Biotechnol. Biochem. 63 (2), 452-455 (1999)  
PUBMED 99209008  
REFERENCE 5 (sites)  
AUTHORS TAKAMI, H., TAKAKI, Y., NAKASONE, K., SAKIYAMA, T., MAENO, G.,  
SASAKI, R., HIRAMA, C., FUJI, F. and MASUI, N.  
TITLE Genetic analysis of the chromosome of alkaliphilic Bacillus  
JOURNAL halodurans C-125  
MEDLINE Extremophiles 3 (3), 227-233 (1999)  
PUBMED 99411980  
REFERENCE 6 (sites)  
AUTHORS TAKAMI, H.  
TITLE Genome analysis of facultatively alkaliphilic Bacillus halodurans  
JOURNAL C-125  
MEDLINE Extremophiles in deep-sea environments, 249-264 (1999)  
PUBMED 7 (sites)  
AUTHORS TAKAMI, H., MASUI, N., NAKASONE, K. and HORIKOSHI, K.  
TITLE Replication origin region of the chromosome of alkaliphilic  
JOURNAL Bacillus halodurans C-125  
MEDLINE Biosci. Biotechnol. Biochem. 63 (6), 1134-1137 (1999)  
PUBMED 99356711  
REFERENCE 8 (sites)  
AUTHORS TAKAMI, H. and HORIKOSHI, K.  
TITLE Analysis of the genome of an alkaliphilic Bacillus strain from an  
JOURNAL industrial point of view  
MEDLINE Extremophiles 4 (2), 99-108 (2000)  
PUBMED 20263314  
REFERENCE 9 (sites)  
AUTHORS NAKASONE, K., MASUI, N., TAKAKI, Y., SASAKI, R., MAENO, G., SAKIYAMA, T.,  
HIRAMA, C., FUJI, F. and TAKAMI, H.  
TITLE Characterization and comparative study of the rrm operons of  
JOURNAL alkaliphilic Bacillus halodurans C-125  
MEDLINE Extremophiles 4 (4), 209-214 (2000)  
PUBMED 20426005  
REFERENCE 10 (sites)  
AUTHORS TAKAMI, H., NAKASONE, K., TAKAKI, Y., MAENO, G., SASAKI, R., MASUI, N.,  
FUJI, F., HIRAMA, C., NAKAMURA, Y., OGASAWARA, N., KUHARA, S. and  
HORIKOSHI, K.  
TITLE Complete genome sequence of the alkaliphilic bacterium Bacillus  
JOURNAL halodurans and genomic sequence comparison with Bacillus subtilis  
MEDLINE Nucleic Acids Res. 28 (21), 4317-4331 (2000)  
PUBMED 20512582  
REFERENCE 11 (bases 1 to 303249)  
AUTHORS TAKAMI, H. and TAKAKI, Y.  
TITLE Direct Submission

JOURNAL Submitted (22-MAR-2000) Hideto Takami, Japan Marine Science and  
Technology Center, Deep-sea Microorganisms Research Group; 2-15  
Natsushima, Yokosuka, Kanagawa 227-0061, Japan  
COMMENT (E-mail: takami@jamstec.go.jp)  
URL: <http://www.jamstec.go.jp/jamstec-e/bio/DEEPSTAR/Research.html>,  
Tel: 81-468-67-3895, Fax: 81-468-66-6364  
On Jan 31, 2001 this sequence version replaced gi:10174886.  
FEATURES  
location/Qualifiers  
1..303249  
    /organism="Bacillus halodurans"  
    /mol\_type="genomic DNA"  
    /db\_xref="taxon:8665"  
    /note="alkaliphile"  
104..1102  
    /gene="BH2267"  
104..1102  
    /gene="BH2267"  
    /codon\_start=1  
    /transl\_table=1  
    /product="penicillin-binding protein"  
    /protein\_id="BAB05986.1"  
    /db\_xref="GI:10174887"  
    /translation="MKVHHTIMEDWLSGTALMKKEDIPFISLGFANPAENPNE  
HHTREGLASGCKLFTALICQLVEAGLSFDTPLSDWLDAPFNVTIHHLLHTSGVP  
DYFDEETDDFEDLMKDVPMYHLRLKDFLPFGHAPKPEPPGHRPHYNNAGFTLLGL  
VVESSGVTFOEYVYANVPORAGWHSYGYPFDLPKATLGYLDLDEGSKTLLYSL  
PVISGSDGAVVTAEDEMMKLMALREHLNETYTKLTPHVAHCEDDDYGYGVNWK  
QQDAISKHYMGVDPVCFHSAFYFSPNSGIVVCAQSSGAYVMAIEALFESEA"  
complement (11146..1469)  
    /gene="BH2268"  
complement (11146..1469)  
    /gene="BH2268"  
    /note="BH2268"  
unknown"  
    /codon\_start=1  
    /transl\_table=1  
    /protein\_id="BAB05987.1"  
    /db\_xref="GI:10174888"  
    /translation="MKVITLGGSTKFKROEQAAYLTIKGHVVLVSGFEPSGRIE  
TEBOVQFENIRKIKMSDEIFVLVDVGYIGKSTREIEFATSKGMPRIKYSSELA  
STFS"  
complement (1636..2373)  
    /gene="BH2269"  
complement (1636..2373)  
    /gene="BH2269"  
    /note="BH2269"  
unknown"  
    /codon\_start=1  
    /transl\_table=1  
    /protein\_id="BAB05988.1"  
    /db\_xref="GI:10174889"  
    /translation="WVKKVGILVGVPLSIGLACNGQOEVEKNTIDGASGIPSI  
EEIIDLNNLPDPAATVNGESIIDEKTYVAVLEQQMTAAOYGISLDEGKQMSLIEE  
QTEIQLNEQLQQAATKEKEIYESEEDIDELAFIVAFQSEELMKALDEGSGSMBE  
LREIEIHYHQQKVEEETEAIVNTEIEIQARVEEKQYVTEDELPTFEEIQSNIIEQR  
LIAKEQEOLEVEPEKLRDGDITVHI"  
complement (2582..3367)  
    /gene="BH2270"  
complement (2582..3367)  
    /gene="BH2270"  
    /codon\_start=1  
    /transl\_table=1  
    /product="ABC transporter (ATP-binding protein)"  
    /protein\_id="BAB05989.1"  
    /db\_xref="GI:10174890"  
    /translation="MADILQEVQTLTKAKKPILOQISWQEGEHNAIVLNGSGKT  
SLIKITIGVMPTEGAVAVGNRGVPIQDEVKRIQVMSLDRFHTGQDVLVI  
VLSGHGTVAVEPEIGADVAKAAEATETGLTHLDERFTLSSGSKKAFARAVY  
AKPEFTLDEPTGGLDIAREOQLTLESVATTKGAPPLVYTHYPEIQPFIHVLM  
LKQGHVIAQGRKQDEVLTENITTAFGCSLDVIAQSGRYWVKPM"  
complement (3378..4310)  
    /gene="BH2271"  
complement (3378..4310)

```
/gene="BH2271"
/codon_start=1
/codon_table=11
/product="dipeptidase"
/protein_id="BAB05990.1"
/db_xref="GI:10174891"
/translation="MRIPAHCDALIKMEDPAREYTDPAIENAVRMEGVAAOC
FAIVGPNIPFDOKFOVAQAQELFEKILKHPQSFTERMDLITLDAQIGALT
LEGADPFDDPTLKTLYRLGVRSGLTNQNLACADGEBRGGLTEWGAVALN
NEHVLTVDVSHLSERGFVIEYAPFVATHSNAKAICHRRLTDAQKALIEAGF
IGIVFHPHFLTGTDDASIDVMKHEICSLGGQKVGSGDFDGI PCFVDGLEHAGR
FQWFMDELKRYDEELVRGFAMDNFLKATPISTV"
4406..4840
/gene="BH2272"
/406..4840
/gene="BH2272"
/codon_start=1
/codon_table=11
/product="heat shock protein class I (low molecular
weight)"
/protein_id="BAB05991.1"
/db_xref="GI:10174892"
/translation="MSRESGSKLPVNYODFLRSIDPFPSKSFQHLNDNGMFAPISIP
IROYKEQPIIEALPVSQKQIQLDVPQYVRIHEEVLMEDEDEKGYXSIS
SOAREVIEPFYPVQEGDIKARLRLNLITAIPIHNRKINIE"
4864..5193
/gene="BH2273"
/4864..5193
/gene="BH2273"
/5193
/unknown
/codon_start=1
/codon_table=11
/protein_id="BAB05992.1"
/db_xref="GI:10174893"
/translation="MYNLHPROTWSQAYYQDPQHSWPNHPQPTSSGMQMYSTG
PSEPFYPNVFSQOKPKASPMKAAFTNODGFINDKTFQYTDQVVKVNYQTPLVKS
VGSIFIK"
complement(5202..5510)
/gene="BH2274"
/complement(5202..5510)
/gene="BH2274"
/5510
/unknown
/notes="BH2274"
/notes="BH2274"

Query Match 10.0%; Score 62.4; DB 1; Length 303249;
Best Local Similarity 49.0%; Pred. No. 0.017;
Matches 196; Conservative 0; Mismatches 201; Indels 3; Gaps 1;
```

```
Db 170831 ATAGATCCAAATTAAAGAGACTTAACGAATTTTCGCTTC 170792

RESULT 5
BSFLHB BSLHB 1105 bp DNA linear BCT 17-MAR-1994
LOCUS BSLHB
DEFINITION B.subtilis flhb gene.
ACCESSION X74121
VERSION X74121.1 GI:395389
KEYWORDS flagellar protein; flhb gene.
SOURCE Bacillus subtilis
ORGANISM Bacillus subtilis
REFERENCE 1
AUTHORS Carpenter, P.B., Zuberi, A.R. and Ordal, G.W.
TITLE Bacillus subtilis flagellar proteins Flp, FlpQ, FlpR and Flhb are
related to Shigella flexneri virulence factors
JOURNAL Gene 137 (2), 243-245 (1993)
MEDLINE 94131291
PUBMED 8299954
REFERENCE 2 (bases 1 to 1105)
AUTHORS Carpenter, P.B.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-1993) P. B. Carpenter, University of Illinois, 190
Medical Sciences Building, 506 S Mathews, Urbana IL 61801, USA
FEATURES
source
1..1105
/organism="Bacillus subtilis"
/mol_type="genomic DNA"
/isolate="O11085"
/db_xref="taxon:1423"
1..1083
/gene="flhb"
1..1083
/gene="flhb"
/codon_start=1
/translation="flagellar biosynthetic protein"
/product="flagellar biosynthetic protein"
/protein_id="CA52218.1"
/db_xref="GI:395390"
/db_xref="SWISS-PROT:P35538"
/translation="MKLRVLDQFAGEKTEKATPKRKRKDTKKGQVAKSDVNSAVSL
LVFLSLAIGPVMRDLSFIEFPEYSLMKSESVMHFLFSLKDMGMLAPIL
LVALVAGVSNYMGVGFSAEIVOPTLEKIDPKPKRIYSMAIYELISIIKIV
VGRAPAVLMLHGEIIRLPILTEBALSTFSKTTMGSLGASALILADILYKOR
FDYENIKMSKQIDIKDEYKSGDPIIKSKIKOROREAMRMQEVPAKADVIITNPT
HYAIALKYEKMDAPYIVAGVDHMLKIRKIAKHEHVMVNRPLARALYDVEID
QAVDEEFKVLAEILAVVYKQKQY"
BASE COUNT 355 a . 192 c 252 g 306 t
ORIGIN

Query Match 9.9%; Score 61.6; DB 1; Length 1105;
Best Local Similarity 47.4%; Pred. No. 0.12;
Matches 184; Conservative 0; Mismatches 204; Indels 0; Gaps 0;
```

Db	Accession	Version	Keywords	Source
Db	280	ATGGGCATGATACATCGAACCCGATCTTGGCTTGTGCTTGTGATGGAGGTGGTAGCAAC	339	
Qy	319	CGTGTACAGTTGGTGCATATATGACATCAAGGTTTTTAATTAAATGAGTAATT	378	
Db	340	TATATGACAGGTGGTGGTTCTTCTTCTCTGAAAGATACAGCCATTAAGGTTGAAACTG	399	
Qy	379	AATTAATAAAGGTTGAAAGGATGT	406	
Db	400	GATCCGATCAACGCTTTAAACGAAATAT	427	
RESULT 6				
BSUB0009				
LOCUS				
DEFINITION	BSUB0009	208780 bp	DNA	linear
ACCESSION				
VERSION	299112	AL009126		
KEYWORDS	299112.1	GI:2633902		
SOURCE				
ORGANISM	Bacillus subtilis subsp. subtilis str. 168			
REFERENCE	Bacillus subtilis subsp. subtilis str. 168			
AUTHORS	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
	1 (bases 1 to 208780)			
	Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azavedo, J., Bertero, M.G., Besieres, P., Bolochin, A., Borcher, S., Borriss, R., Bouvier, L., Brans, A., Braun, M., Brignell, S.C., Bron, S., Broiliet, S., Bruschi, C.V., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Conerton, I.F., Cummings, N.J., Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D., Emerson, P.T., Enright, K.D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Funa, S., Galizzi, A., Gallerton, N., Gilm, S.Y., Glaser, P., Goffeau, A., Goldlight, E.J., Grandi, G., Giuseppe, G., Guy, B., J., Hage, K., Haech, J., Harwood, C.R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F., Ileya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klaer, Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningsstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H., Maundu, M., Manuel, C., Medigue, C., Medina, N., Mellado, R.P., Minano, M., Moesti, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogasawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M., Portetelle, D., Porwollik, S., Prescott, A.M., Presneau, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivola, C., Roche, E., Roche, B., Rose, M., Sadate, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, D., Sekowka, A., Seror, S.J., Serro, P., Shin, B.S., Soldo, B., Sotokita, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpetra, P., Tognoni, A., Toso, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassartoli, A., Vairi, A., Wambuit, R., Wedler, E., Wedler, H., Weltzinger, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H.F., Zumbstein, E., Yoshikawa, H. and Danchin, A.			
TITLE	The complete genome sequence of the gram-positive bacterium			
JOURNAL	Bacillus subtilis			
MEDLINE	Nature 390 (6657), 249-256 (1997)			
PUBMED	98044033			
AUTHORS	2 (bases 1 to 208780)			
TITLE	Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.			
JOURNAL	Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur, Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr.			
FEATURES	adanchin@pasteur.fr Phone: +33 (0) 1 45 68 84 41, Fax: +33 (0) 1 45 68 89 48			
source	Location/Qualifiers			
	1. 208780			
	/organism="Bacillus subtilis subsp. subtilis str. 168"			
	/mol_type="genomic DNA"			
	/strain="168"			
	/db_xref="taxon:224308"			

gene .4588  
/gene="bpr"  
258..4559  
/gene="bpr"  
/EC number="3.4.21.-"  
/note="alternate gene name: bpf"  
/codon start=1  
/transl\_table=11  
/product="bacillopeptidase F"  
/protein\_id="CAB13404.1"  
/db\_xref="GI:2633903"  
/db\_xref="GOA:P16397"  
/db\_xref="SWISS-PROT:P16397"  
/translation="MKRTKRLISLVLTIVTISLLPFGAAGSSKVTSPSKLEQ  
SNEISQNTISSLSKFSFKKKEKTFLLIKFKDLANPEKAAGA VKAKSKLSTAKLEY  
OKRSEAVSLKATLDESQQDLKYNTQKRGADQIHYYVNGLAIVASKEWKEV  
VQPEKVVLTLENKROLDFKSSSPNNKKAQAKIATQIGVMYNDQIDA PAMALGDXG  
TGTVASIDTQGENMHPALKEKRYGVNDPENDEBNMNYDVAAGASPYDLAQT  
HYTMGASPEPDGTNOICVARGAKWIVKAPSEDCGDDNII LAGEWVILA PKAENP  
HPBNAPDVYNNSSKGGSGLDWRYEDMYNMRADIPFEPSAGNTDLFI PGCPSTANP  
ANYPEPSATGATDINKKLADSLQSPSPYEIKETISAPGVNIRSSVGGYEDWDG  
TSMAPPHSAVAALIKQANMLSYDEMEDILITSTAEPLSTDPDSFNNQYGHGVNA  
FDVAEAVDGGKAEQGVSEGDQEPVYQHEKVTAYAGSGSLPLTLMEDVYNSD  
VKLSKILQDGEWTELTAKRISGHLKGTVOAEPIDIGTGLSYKMMHHDGHHVSSD  
VYDVYKRSITAGYKQDEPTAPGCVASGTRNNMKNVPSPTGPTASGKRYGNTLT  
GNYASAMNMLVMPPIKAPDSGS.LQPKSWHNLEDPDFDYGTVFLPEGSKMWEQGV  
NGKTSITWDEIDLSAKGQNIQVMENLOSDEIAKEGVYIDVVLSDSACKTYK  
NGLGTERKSGKQKVPVPPKAKPSANTAVHOKRALQPOLPLKQVSVETGKSTY  
SDQSGQVTLGHAKGADYTLMAEAVGYSTOKVSLKLTDOITQANFTLEMKKGLKGT  
VINKTSGKVAAGVYVEDAAVEAPMTNDKGEVMLEAYEAGVATIKVAAGYSDERS  
VELKQDYKERTALRPFVQYRBEIAYDGTANANSYTAAGNGAVKRTIADGKDKML  
TGGLRFDITFEPDPGTEFEVEYIDA TGDGAGKTIAGPUNAALRNDEWTKVDLS  
SKGIWVADFLVYIQSKPDPYSGMLADEMGQSGNNWQYIDGKWPGBKADGNTMI  
BALVYVADFAVEIITSPTDKSYTKDSYTVGNGASPGTTHIYNGEKAGETKAAADST  
FHAGIILKNGENMLPATSTNGTGTDSAPITVTLDOKEBELTLDNPKDGKTKETV  
TYKGVSDNDIKDVKNGKKATVADGSYARILLENERNIKIYATIDAGNKTKTKTV  
IDVNDKRVISGLIPGEKNTLKAGESYKIASSEMEDDAFTTIRMPITNARAYQNT  
ELPLEISPGRYEGWTATSSIRAKAKAVEIYVADYGNETRTKANGLNMNTEN"

terminator .4588  
/gene="bpr"  
4754..5683  
/gene="spoilGA"  
4754..5683  
/gene="spoilGA"  
/function="processing of pro-sigma-E (SpoIIGB) to active  
sigma-E (stage II sporulation)"  
/codon start=1  
/transl\_table=11  
/product="protease"  
/protein\_id="CAB13405.1"  
/db\_xref="GI:2633904"  
/db\_xref="GOA:P13801"  
/db\_xref="SWISS-PROT:P13801"  
/translation="WKIYDVYIMLNFCSFDALLLTAFLKXHYKRRALVGAFTGS  
SIVALLPFPSPVIEHPAGKLAFSVIVAVTFGKGRFFQULFSFYFTFPMGGI  
IGASHLLOSNISVONGVMTNQTFGRPISMLPIVGFSPALMFEFSKRIDITPKQI  
YEERVAIVDAGSQTILHGRGIDSGNQLYDPLTQPTVMIYIDKLEIIFSTALTIMIR  
NTDEPLAEIADLDSFRFLDKMRILPYRGVQGNQFLLCVAPDHVITINTKEMISADKC  
LIGISTTSLSDGEFDALIHKMLSGAAVHVS"  
5746..6465  
/gene="sigE"  
/gene="sigE"  
5746...6465  
/gene="sigE"  
/function="early mother cell-specific gene expression"  
/note="alternate gene name: spoIIGB"  
/codon start=1  
/transl\_table=11  
/product="RNA polymerase sporulation-specific sigma-29  
factor (sigma-E)"  
/protein\_id="CAB13406.1"  
/db\_xref="GI:2633905"  
/db\_xref="GOA:P06222"  
/db\_xref="SWISS-PROT:P06222"

gene .4599  
/gene="bpr"  
258..4559  
/gene="bpr"  
/EC number="3.4.21.-"  
/note="alternate gene name: bpf"  
/codon start=1  
/transl\_table=11  
/product="bacillopeptidase F"  
/protein\_id="CAB13404.1"  
/db\_xref="GI:2633903"  
/db\_xref="GOA:P16397"  
/db\_xref="SWISS-PROT:P16397"  
/translation="MKRTKRLISLVLTIVTISLLPFGAAGSSKVTSPSKLEQ  
SNEISQNTISSLSKFSFKKKEKTFLLIKFKDLANPEKAAGA VKAKSKLSTAKLEY  
OKRSEAVSLKATLDESQQDLKYNTQKRGADQIHYYVNGLAIVASKEWKEV  
VQPEKVVLTLENKROLDFKSSSPNNKKAQAKIATQIGVMYNDQIDA PAMALGDXG  
TGTVASIDTQGENMHPALKEKRYGVNDPENDEBNMNYDVAAGASPYDLAQT  
HYTMGASPEPDGTNOICVARGAKWIVKAPSEDCGDDNII LAGEWVILA PKAENP  
HPBNAPDVYNNSSKGGSGLDWRYEDMYNMRADIPFEPSAGNTDLFI PGCPSTANP  
ANYPEPSATGATDINKKLADSLQSPSPYEIKETISAPGVNIRSSVGGYEDWDG  
TSMAPPHSAVAALIKQANMLSYDEMEDILITSTAEPLSTDPDSFNNQYGHGVNA  
FDVAEAVDGGKAEQGVSEGDQEPVYQHEKVTAYAGSGSLPLTLMEDVYNSD  
VKLSKILQDGEWTELTAKRISGHLKGTVOAEPIDIGTGLSYKMMHHDGHHVSSD  
VYDVYKRSITAGYKQDEPTAPGCVASGTRNNMKNVPSPTGPTASGKRYGNTLT  
GNYASAMNMLVMPPIKAPDSGS.LQPKSWHNLEDPDFDYGTVFLPEGSKMWEQGV  
NGKTSITWDEIDLSAKGQNIQVMENLOSDEIAKEGVYIDVVLSDSACKTYK  
NGLGTERKSGKQKVPVPPKAKPSANTAVHOKRALQPOLPLKQVSVETGKSTY  
SDQSGQVTLGHAKGADYTLMAEAVGYSTOKVSLKLTDOITQANFTLEMKKGLKGT  
VINKTSGKVAAGVYVEDAAVEAPMTNDKGEVMLEAYEAGVATIKVAAGYSDERS  
VELKQDYKERTALRPFVQYRBEIAYDGTANANSYTAAGNGAVKRTIADGKDKML  
TGGLRFDITFEPDPGTEFEVEYIDA TGDGAGKTIAGPUNAALRNDEWTKVDLS  
SKGIWVADFLVYIQSKPDPYSGMLADEMGQSGNNWQYIDGKWPGBKADGNTMI  
BALVYVADFAVEIITSPTDKSYTKDSYTVGNGASPGTTHIYNGEKAGETKAAADST  
FHAGIILKNGENMLPATSTNGTGTDSAPITVTLDOKEBELTLDNPKDGKTKETV  
TYKGVSDNDIKDVKNGKKATVADGSYARILLENERNIKIYATIDAGNKTKTKTV  
IDVNDKRVISGLIPGEKNTLKAGESYKIASSEMEDDAFTTIRMPITNARAYQNT  
ELPLEISPGRYEGWTATSSIRAKAKAVEIYVADYGNETRTKANGLNMNTEN"

terminator .4599  
/gene="bpr"  
4754..5683  
/gene="spoilGA"  
4754..5683  
/gene="spoilGA"  
/function="processing of pro-sigma-E (SpoIIGB) to active  
sigma-E (stage II sporulation)"  
/codon start=1  
/transl\_table=11  
/product="protease"  
/protein\_id="CAB13405.1"  
/db\_xref="GI:2633904"  
/db\_xref="GOA:P13801"  
/db\_xref="SWISS-PROT:P13801"  
/translation="WKIYDVYIMLNFCSFDALLLTAFLKXHYKRRALVGAFTGS  
SIVALLPFPSPVIEHPAGKLAFSVIVAVTFGKGRFFQULFSFYFTFPMGGI  
IGASHLLOSNISVONGVMTNQTFGRPISMLPIVGFSPALMFEFSKRIDITPKQI  
YEERVAIVDAGSQTILHGRGIDSGNQLYDPLTQPTVMIYIDKLEIIFSTALTIMIR  
NTDEPLAEIADLDSFRFLDKMRILPYRGVQGNQFLLCVAPDHVITINTKEMISADKC  
LIGISTTSLSDGEFDALIHKMLSGAAVHVS"  
5746..6465  
/gene="sigE"  
/gene="sigE"  
5746...6465  
/gene="sigE"  
/function="early mother cell-specific gene expression"  
/note="alternate gene name: spoIIGB"  
/codon start=1  
/transl\_table=11  
/product="RNA polymerase sporulation-specific sigma-29  
factor (sigma-E)"  
/protein\_id="CAB13406.1"  
/db\_xref="GI:2633905"  
/db\_xref="GOA:P06222"  
/db\_xref="SWISS-PROT:P06222"

/translation="MKKRLRLTHIMYKLMKLGKSDYVYIGSSSLPPPLSKDEE  
OVLMLKPNQDAPARAILITERILRLVITAKFENGITNIEDLSIGITGLIKAVNT  
NPKKIKLAVASRCIENEILMYLRNKRKISEVSFYPPLNDMGNELLSIGITGLI  
DDITKDIENAVDSKLLKALQOLNERKQIMELFGLVGESEKTKOXVADMGISOS  
YISLRKRIIKRLRKEFNKNV"  
6605..7387  
/gene="sigc"  
6605..7387  
/gene="sigc"  
/function="late forespore-specific gene expression"  
/note="alternate gene name: spoIIIG"  
/codon start=1  
/transl\_table=11  
/product="RNA polymerase sporulation-specific sigma factor  
('sigma-G')"  
/protein\_id="CAB13407.1"  
/db\_xref="GI:2633906"  
/db\_xref="GOA:P19940"  
/db\_xref="SMIS-PROT:P19940"  
/translation="MSRNKVEICGYDTSKLYLKREMRKLFROLQDEGDSAREKLV  
NGNKLVLVITQRFNRRGSEVYDDLFQVSCIGLMSIDNFDLSHNVKSTVAVPMITGE  
IRRYLRDNPRLRVSRSLDIKALQVBERLISSTSKPELADIARVLEHBEIVFA  
LDAIODPVSLPEPILYVDGDIYVMDQISDERNTDSQWIEELALKEGRRLRINDEKMI  
LRKRFQCKTQMEVAEIEIGISOQVSRLEKAIKQMNKHQ"  
7535..8496  
/gene="yIma"  
7535..8329  
/gene="yIma"  
/function="unknown"  
/note="similar to ABC transporter (ATP-binding protein)"  
/codon start=1  
/transl\_table=11  
/protein\_id="CAB13408.1"  
/db\_xref="GI:2633907"  
/db\_xref="GOA:O31723"  
/db\_xref="SPTREMBL:O31723"  
/translation="MILQDNVSLKRNKWLKIDHMYKEKNVLYGLNGAGKTAL  
LNMCSYFPTSGEKRVLGHEGKTELEGKRLKIGVSAALQOLYPADARAFETAP  
GAVASIGVETPSKRETKAIGLEDIGATVAPRVEYTSQGEKOPALIRALMADP  
ELLIDPEVTCIDPTAREKLDITTYTANKENAPSLVYTHAEELIPVFKALLKQ  
GEVFGSGIKEMLTDIILSAFFDPIHVLNMDRFLTRAEPITVA"  
8472..8496  
/gene="yIma"  
8531..9811  
/gene="yImb"  
8531..9811  
/gene="yImb"  
/function="unknown"  
/note="similar to acetylornithine deacetylase"  
/codon start=1  
/transl\_table=11  
/protein\_id="CAB13409.1"  
/db\_xref="GI:2633908"  
/db\_xref="GOA:O31724"  
/db\_xref="SPTREMBL:O31724"  
/translation="MDQIYISLOKKEHKEKELIOLAKTLISYOTPARPANTGESIOS  
WLAGYINELGSIKMDYVPGDPRNVGSLKATGSDVGLIINGHVDLVAIKYEBEMK  
HDPRIPIREKGLIGRGSMDKMGACVLPVAKLRKESIELPDLIIQSYIGEVEGE  
AGTLECCRGHADPAIVADTSDMHIQOQGVITGWIIEIKSQTEHDTGRRAHAG  
GTFGASALEKAKIAGLEBERHMSYKPGKPTNPVATIEGRRAHAFAD  
CLMTVHVPNEHTDQVAIEEDVNSDIDILRENRFVFKGSGSMEDGEIIF  
PALEVDPGHPGIALTASHOKVKECPDIIDVOSVYDGMVYDAGIPCVIAPGLHN  
AHSVNEKYSIOIVAYETKIILDFIISWRKKEQ"  
9894..10139  
/gene="yImc"  
9894..10139  
/gene="yImc"  
/function="unknown"

Query Match 9.9%; Score 61.6; DB 1; Length 208780;  
Best Local Similarity 47.4%; Pred. No. 0.027; Indels 0; Gaps 0;  
Matches 184; Conservative 0; Mismatches 204; Indels 0; Gaps 0;

Qy 19 AAACAGAGAAAGCAACCCCGAAACGACGTACAGAGCTGCTTGAAGGAGTCCCT 78  
Db 107613 AAGACAGAAAAAGCACCTCCGAAAAAAGCAGAGATACGCAAAAAAGGCGAGGTAGCC 107672  
Qy 79 AAATCAGAGAGGTACTTAAGACATGACCTACACAGAGGATGCTGGGCGTCTAT 138  
Db 107673 AAAAGCTCGATGTCATACATCCGCGTTCCTTACTGGAATATTCCTTTCGCTATTGCA 107732  
Qy 139 TATTCAGCGCTAATAGGACGTCACTTTTGAACAAATTTTCATATTTTACAGATCA 198  
Db 107733 ATTGCGCGATATGAGACAGAGCGCTGCTCATTTACAGAAATTTTATACCGAATCG 107792  
Qy 139 TTTGCGTTTGAAGTTACAGACACTGACGTATATTCCTTATTTATTTGTTGCTCAAG 258  
Db 107793 CTCAATAGAGCTCTCAGATCAATGTGATACATTTGTTGACGCTTATTAAGGAT 107852  
Qy 259 ATATCTATTTATTTATGATGCAATATTTATTTTATGCTGTGATGAGCATTTTCAATTA 318  
Db 107853 ATGGGCAATGATATCTCGACCGATCTTGTCTTCTGTTGAGAGAGTGTGAAGCAAC 107912  
Qy 319 CGTGTACAGAGTGTGATTAATGACATCAAGGTTTAAATTTAAATGAGTAAATTT 378  
Db 107913 TATATGAGAGTGTGGTCTTCTTCTTCTGCTGAAGTATACAGCTTGAAGAACTG 107972  
Qy 379 AATATATTAAGAGGTTGAAGAGATGT 406  
Db 107973 GATCCGATCAAGGCTTAAACGATAT 108000

# RESULT 7 PFMALBP1\_01/c WPCOMMENT

Sequence split into 14 fragments LOCUS PFMALBP1 Accession AL844507

Fragment Name	Begin	End
PFMALBP1_00	1	110000
PFMALBP1_01	100001	210000
PFMALBP1_02	200001	310000
PFMALBP1_03	300001	410000
PFMALBP1_04	400001	510000
PFMALBP1_05	500001	610000
PFMALBP1_06	600001	710000
PFMALBP1_07	700001	810000
PFMALBP1_08	800001	910000
PFMALBP1_09	900001	1010000
PFMALBP1_10	1000001	1110000
PFMALBP1_11	1100001	1210000
PFMALBP1_12	1200001	1310000
PFMALBP1_13	1300001	1325595

Continuation (2 of 14) of PFMALBP1 from base 100001 (AL844507 Plasmidium falciparum 3D7)

Query Match 9.3%; Score 57.8; DB 2; Length 110000;  
Best Local Similarity 45.8%; Pred. No. 0.16; Indels 0; Gaps 0;  
Matches 200; Conservative 0; Mismatches 237; Indels 0; Gaps 0;

Qy 173 TTTTCTACTATATTTTACAGATCATTTGCGTTGAGGTTACAGACAGTCAGTATXTG 232  
Db 26171 TTGTTTTTTTTTTGTTTGTGTTATGTTTTTTTTTTTCTTTTGATATGTCATTTGAT 26112  
Qy 233 CTTTATTTATTTATTTGTTGCTCAGAGATGCTATTTATTTAGTCCATATTTACTTTTA 292  
Db 26111 TTTTGCTGTTTCTTCTACAGCTATTTTGTGATAAAGTATTGCAATTTATTTTGA 26052  
Qy 293 TTGCTGTACGCGATGATTTTCATTTAGTGACAGTGTGGCATTTTGTGCTACAAAG 352  
Db 26051 ATTTGTTATCTTTTGTCTCAGCTATTTAAGTGTGTTTCTTATTTATTTATTTG 25992  
Qy 353 TTTTAAATTTAAATGAGTAAATTTATATTAATAAAGGTTGAAGGATGTTGCTT 412  
Db 25991 TTTTATAGTTGATACCTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 25932  
Qy 413 CTCAACAACACTGTTGACCTTTTACGATTTTATTTCAAGTATTTATTTATTTATTTG 472  
Db 25931 TGCCATAATTTATTTGTTGAAATTTATTTATTTATTTATTTATTTATTTATTTATTTG 25872





```

/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT      3711 a      371 c      371 g      3603 t
ORIGIN

```

Query Match	8.8%	Score 54.8	DB 6	Length 8056
Best Local Similarity	47.2%	Pred. No. 1.1		
Matches 167; Conservative	0	Mismatches 187	Indels 0	Gaps 0

Qy	160	CATTGGAAACATTTCTACATATTTTTCAGACATCATTCGGTTGAGGTACAGA	219
Db	6351	CGTAAATAATAATTTTTATATATTTATTAATTAATTAATTAATTAATTAAT	6410
Qy	220	CAGTCAGATATGCTTATTTATTTATGCTCAAGATAGCATTTTATATGATGCA	279
Db	6411	ATATATTTATTTAAATTTAAATTTTTTATTTAAATTTAAATTTTTTATTTATTA	6470
Qy	280	ATATTACTTTTATTTATGCTTACGGCATGGAATTCATTACGTGACAAAGTTGGTCATTA	339
Db	6471	ATTAATAATATCGTTTTTTTATTTTCGTTATTTAAATTTTCGTAATTTTCGACGAA	6530
Qy	340	TGACACTCAAAAGTTTTTAATTTAATTAATGAGCTAATTTATATATAATAAAGCGTTGAA	399
Db	6531	AACGTTTATTTAATTTAAATAATAATTTATTTAATTTAAATTTAATTTAAACGAAAAA	6590
Qy	400	GGAATGTTGCTCTCAACAAACACTGTGCACTTTACGACGTTTACGTAATAT	459
Db	6591	AAATATTTATTTTATTTAATTTAATTTTAATTTTAATTTTATTTTACGTTTTTTATACGA	6650
Qy	460	GTTATAGGATTTGTCATATATGATTTAATAAAGAGATTTTCAACCTTTTA	513
Db	6651	AAATTTTATTTTATTTAATTAATTTTAAATAAATAATTTTATTTTACGTTTA	6704

FEATURES	source
LOCUS	AX345047
DEFINITION	Sequence 118 from Patent WO0200928.
ACCESSION	AX345047
VERSION	AX345047.1
KEYWORDS	GI:18492933
SOURCE	synthetic construct
ORGANISM	synthetic construct
REFERENCE	artificial sequences.
AUTHORS	1
TITLE	Olef, A., Piepenbrock, C. and Berlin, K.
JOURNAL	Diagnosis of diseases associated with the immune system
	Patent: WO 0200928-A 118 03-JAN-2002;
	Epigenomics AG (DE)
	Location/Qualifiers
	1..9052
	/organism="synthetic construct"
	/mol_type="genomic DNA"
	/db_xref="taxon:32630"
	/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT	2166 a 112 C 2104 g 4670 t
ORIGIN	

Query Match	8.8%	Score 54.8	DB 6	Length 9052
Best Local Similarity	47.2%	Pred. No. 1.1		
Matches	167	Conservative	0	Mismatches 187, Indels 0, Gaps 0
Qy	138	TTATTACAGCGCGAATGGGACGTCATTGTTGAAACAAATTTCTACTATATTTTACAGAAATC	197	
Db	799	TTTTTTAGATATAAAAGAGTTAGTTATTTAGTTTGGTTTATTTATTTATTTGGTAT	858	
Qy	198	ATTTCGGTTTGAGGTTACAGACAGTCAGATATGCTTTATTTATTTATGTTGCTCAAGA	257	
Db	859	TTATTTTTTTTATTTATTTTATTTATTTGTTATTTATTTTATTTTACGATATATATTA	918	
Qy	258	GATAGCTATTTTATTTAGTGCACAAATTTACTTTTATTTGCTGTACGCAATGATTTTCAT	317	

[illegible]

RESULT	11
LOCUS	AC116967
DEFINITION	19237 bp DNA linear INV 12-MAR-2003
ACCESSION	AC116967
VERSION	AC116967.2
KEYWORDS	AK4, complete sequence.
SOURCE	HTG.
ORGANISM	Dicystostelium discoideum
REFERENCE	Dicystostelium discoideum
AUTHORS	Eukaryota; Mycetozoa; Dictyostelid; Dictyostelium. 1 (bases 1 to 19237) Glöckner,G., Eichinger,L., Szafrański,K., Pachebat,J., Dear,P., Leimann,R., Baumgart,C., Parra,G., April,J.F., Galigo,R., Kumpf,K., Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and Neugebauer,A.
TITLE	Sequence and analysis of chromosome 2 of Dictyostelium discoideum
JOURNAL	Nature 418 (6893), 79-85 (2002)
MEDLINE	22092622
PUBMED	12097910
REMARK	The Dictyostelium Genome Sequencing Consortium
REFERENCE	2 (bases 1 to 19237)
AUTHORS	Baumgart,C.
TITLE	Direct Submission
JOURNAL	Submitted (04-APR-2002) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
REFERENCE	3 (bases 1 to 19237)
AUTHORS	Baumgart,C.
TITLE	Direct Submission
JOURNAL	Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
REFERENCE	4 (bases 1 to 19237)
AUTHORS	Baumgart,C.
TITLE	Direct Submission
JOURNAL	Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
COMMENT	On Mar 4, 2003 this sequence version replaced gi:1920066. CDS predictions from Genaid do not necessarily reflect true genes. Further information is available from IMB Jena, Department of Genome Analysis (http://genome.imb-jena.de/dictyostelium/) and the University Cologne, Institute for Biochemistry I (http://www.uni-koeln.de/dictyostelium/project.shtml).
FEATURES	Funding Agency : Deutsche Forschungsgemeinschaft (DFG). Location/Qualifiers 1..19237 /organism="Dictyostelium discoideum" /mol_type="genomic DNA" /strain="AX4" /db_xref="taxon:44689" /chromosome="2" /map="5481549-5500786" complement(join(160..641,723..790)) /note="Genaid exon scores (in order of location ranges): 22-51' 0.50 - GSCJ_ID dd_03333" /codon_start=1
CDS	.

/product="hypothetical protein"  
 /protein\_id="AA051103.1"  
 /db\_xref="GI:28828494"  
 /translation="MDELKRTKTNQINALINHHFSNEIDTFYFGDIENELVKEFIEEYF  
 HEGEPFKIVKPYFNHFIIDINDINKIKISLKSNIISNDNIIVPLSKSDAYFINDYVY  
 KYTESINEIFEMALENGALFGYKININDNNNNKNNGNVIVDENEEKINSNNLV  
 CNYIKSDKSSNLTHTLPEFRCK"  
 2131. 2736  
 /note="geneid exon scores (in order of location ranges):  
 50.44 - GSCJ\_ID dd\_01373"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="AA051097.1"  
 /db\_xref="GI:28828488"  
 /translation="MATIFDILNTLNNNNFNKRCQCPINKSNKTIIDILPSMDVTWT  
 NDKLIEFLAGISKDQIDEIDKSDILITIQEGKKLNKQQOQVSEKSSYSPSTLSS  
 KEDSIEEFEDFDIKPSKATSTTTTTEKKKQKDEKTSPPKFFISERSFGNFRKYIDL  
 LKRVAYOLDINSINQFENGGLTITINKKHAHNTIKININ"  
 complement(join(12919..3600,3675..4065))  
 /note="geneid exon scores (in order of location ranges):  
 29.77, 4.69 - GSCJ\_ID dd\_01077"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="AA051098.1"  
 /db\_xref="GI:28828489"  
 /translation="MLKRNKKSLILHIKDKSPFKTKGPPQDFLENOIKIGLYEEF  
 FIDNDSPEVAVISILITNFNFKLFTLNENENLDFELVKNVAKISIDMKCEYFNSFAIEM  
 BELKRTKINDALINHHFSNEINTFYPDIDNPILVKEFIEKYPOEGEVATPPFNHY  
 IDINDINKDIKILISKINSIISNDNIIVPLSKSDAYFINDYVYKYTESEINEIFMAIE  
 NLAGYKIVNNNNNNNNNNKNNKNNKNKIYVDPDENDEKKVKNSDLVCMNFIKSGK  
 KSNLYTIPERCKRVALKVRAKKLTLSIENNIIPLCSTSHDNINSISLIEKDFKYS  
 YSMANVAVSKNKINFE"  
 join(5458..5626,5684..6099)  
 /note="geneid exon scores (in order of location ranges):  
 4.76, 32.09 - GSCJ\_ID dd\_01076"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="AA051099.1"  
 /db\_xref="GI:28828490"  
 /translation="MATIFDILNTNNNNNNFNCKRCQSTNKSNTIIDLIPSDVDTM  
 TNDKLIIEFLAGEKKKLNKQQOQOQKQVLEKTSSTTSLDKREKDSIEEED  
 DIKPSKSDNTVYATTTTKENKEDKSTDKPFIERSFGNFRKYIDLTITLYQ  
 LDINSINTQFERGLTITINKKHAHNTIKININ"  
 complement(join(6342..6990,7089..7459))  
 /note="geneid exon scores (in order of location ranges):  
 31.77, 5.21 - GSCJ\_ID dd\_01075"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="AA051100.1"  
 /db\_xref="GI:28828491"  
 /translation="MLKENHKKSLIPIKDKSFFKTKGPPQDFLENOIKIGLYEY  
 EFDNDSVAVISILITNFNFKLFTLNENENLDFELVKNVAKISIDMKEGYNSIAIME  
 BELKRTKINDALINHHFSNEINTFYPDIDNPILVKEFIEKYPOEGEVATPPFNHY  
 IDINDINKRTVISEKANSIISDNDNIIVPLSKSDAYIIDNYRKATESINEILMAIE  
 NLAGYKIVNNNNNNKVRVVDENENNTINNYNNNVLVCNFIKSGKSNLITLPEE  
 FRCKIATKVRKAKLTLSIENNIIPLCSTADHTKNSLIEKGFKYSYSAVAVTSK  
 NKINE"  
 join(8484..8774,8822..9133)  
 /note="geneid exon scores (in order of location ranges):  
 9.50, 12.11 - GSCJ\_ID dd\_01074"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="AA051101.1"  
 /db\_xref="GI:28828492"  
 /translation="MATIFDILNTLNNNNFNCKRCQSTNKSNTIIDLIPMDVTT  
 NDKLIEFLAGISKDQIDEIDKSDILITIQERKOLNKHNNNNNNNNNNNLNPKHQ  
 LKVEFDDIIPKPSKSDNRTVSTTTATKKNKDEKTSPPKFFISERSFGNFRKYIDL  
 TKRVAYOLDINSINQFENGGLTITINKKHAHNTIKININ"  
 complement(join(9436..11722,11853..12590,12687..12818,  
 12894..12974,13087..13307))  
 /note="geneid exon scores (in order of location ranges):  
 170.09, 23.78, 2.31, 7.81, 7.70 - GSCJ\_ID dd\_01073"  
 /codon\_start=1

```

/product="hypothetical protein"
/protein_id="AA051102.1"
/db_xref="GI:28828493"
/translation="MISLKEFLILFLINFIQGLSVNPIPSSEYNCISTLIQDYNLKF
PKRDNNQYVDCITANTNDIVCTNNINIKITLAINSGSGTNDILDPNVNVCPOWDELI
VKNLQTPNPIFYKKPRAATVPAVYNSISGLIGCTQPLPAKRLIELPSDANTIFRQ
YIDKVTYISLAMSFSDSYYLTLDNDGANSDBSITITRILINSISLSSYKRLQVTEK
ISPRDYSTISFLKQLQITDOSGMLTYITFENQANLIDQYIPNSIIPYIVSAIFPK
PEPEFDITLFEKNHYKYSITLVPGSTFNINGSPLIPAKNIEISFIFGGLYFENPLNF
NNDITLTFNPNFSIKSESPPEPLSGSGSFDFTGTFKFTGLTIDKSMCTEIVASLGLTGL
VSGCACYENNTVNTYSYSLTKQSMFSRASNSESNMPTPTCTTPAKKELINDGIFP
NKVMTYAGNDIGFDVTLTKINLTISLDPITYIVNQNGKNTAFATVAGANNLKNVYFV
LFTTNVNTVYTFPTNINIPPIIKSVLISKTGNETIQVSHTFSSLYLGTDNQVQNG
NCSITTSDFPGAQCTMESGKQLPIANSITPLQIIEIKDSEFNVAIKIKIESNDNQ
VCPNDICSTSVGNVSIICDSSGTCKLPISGICLIEICVDPGSGNGYCNITTEGCI
CDSHSGSDCSLIFIECDPLDCNSNGNVCOTTKGCKEKNMSGPTCLIPHYITSTP
PSTNGVASPFIQFGPDHNNLSITIGNPICPLLNTSVTLNCTAPAPGKVLVYVTO
NNITYSYDKYQTLNINNOTCPKNCNSQGTCTNLGGCKCKNGNGADCSGITNDGNG
NGGSDSDGNGNGNPETDVGDPGIGTNTTISIQEVOFQIFPSLYIEDNGNIVAY
SIRNMTWNTSNGNSKNEIYTLTQIIOGNSIISKIEITNRNKEVATFEADPNLQVQ
SIKISGIIYNNYSKNNLNTLKLIELVSVQIIOISDDECHMDISITINIKKDGSTFN
YIKIKSKNNKLIEGRFINKLIVSDGRPLYLTITSMKNECSDISLDPHICESCIIIDD
PSVLISPEFKNECGNSKDGKRSYVIPAVAVASVGVSAIGALGFLIRKKGESVTLNK
LAKIKK"
complement(join(13861..17107,17185..17904,18265..18346))
/note="Geneid exon scores (in order of location ranges):
246.52, 30.89, 0.31 - GSCC_id dd_01072"
/codon_start=3
/product="similar to Podocoryne carnea. EGF-like protein
(Fragment)"
/protein_id="AA051104.1"
/db_xref="GI:28828495"
/translation="WFLIPGGGTNNILPESNFSTCEPTLWVETLTSHSNDKLTIRLEHI
NSGINLIDISPNVAVVLYLNSDGNAPSGTSLPKNFYNNHNVISSEYKNNINLQSL
LEPSTINPMKTISSNQSLSYALQINQYLPDNSEYKRFYLSATIDAPETRYIDMSF
VNHITATLTFPENGILYHNLGALPILIPESCDNFSLSGNLTVPENNLNMGYSVSEFN
SIKNQFPTKFLSGSGSYFYTNSPNLFEVTVNNGSCVTELSVPTGNLTGLIPSCFACY
FNVTIINSYSKLTQKSMLETRPFGNKFENNTATPTCTTPAKYELINDGIEPKMTATN
DIGFDVTLTKINLTISLDPITYIVNQNGKNTAFATVAGANNLKNVYFVPLFTPNVY
TYTPTNNINPIIKSVLISKTGNETIQVSHTFSSLYLGTDNQVQNGYCNITTEGCI
DPSFAQCTMESGKQLPIANSITPLQIIEIKDSEFNVAIKIKIESNDNQVNCNDCT
SVGNVSIICDSSGACKCLPGEVSGDCLIEGCVDPDGSNGHCDYITTEGCIENSQDS
DCLPLIPCKPTGCGSDSHQNDGSLPLEICPISGLPCNGGLANTCNNOGTGCGDSS
GMLNTAGNAGDSCGLIIPGCGNGGNSNSDGGANNPNPTDGVGDGTGNTISNOEV
OQPLPEKSLYEIDPENGNTVATYLSLONSMTNSPTDENEIITYLSQTIQNSKISIKTE
EITNRNKEVATFEADPNLQVQSIKISGIIYNNYSKNNLNTLKLIELVSVQIIOISDN
ECNEDWTSINTINNKDSEFTENYIKISKNKLIEGFLYKINLVSDGRPLYLTSMKES
KDSISLDPHICESCIIIDDPSVLISPEFKNECGNSKDGKRSYVIPAVAVASVGV
SAIGALGFLIRKKGESVTLNK"
BASE COUNT      7622 a      2154 c      2055 g      7406 t
Query Match      8.8%; Score 54.6; DB 3; Length 19237;
Best Local Similarity 47.1%; Pred. No. 0.98;
Matches 168; Conservative 0; Mismatches 189; Indels 0; Gaps 0;

```

Db 4138 ATTTTATTAATAATCTCTATTAATAAATCAACCTTAATTTAATGATACCATG 4197

Qy 352 GTTTTAAATTTAAATGAGTAATTTATTAATTAATTAAGGTTGAAGAGATTTTCT 411

Db 4198 TTTTAAATTAATTAATAAATAAATTAATTAATAAATTAATTAATTAATTAAT 4257

Qy 412 TCTCAACAAACATCTGTCGACTTTTACGTAAGTTTATGTTCAAGTAATTTAGTAT 471

Db 4258 AATATAATAACACGAAAAATCTAATTAATTAATAAATTAATAAAGAAATTAATTAATA 4317

Qy 472 GTTCATATATGATATTAATAAGAGAGTTTCAACCTTTTACATTAATTAATGCA 528

Db 4318 TTACCAACCAATGATTAAGAACCAATGATTAATTTCTTTTAAACATTTTATGCA 4374

RESULT 12  
AX599046 8056 bp DNA linear PAT 14-FEB-2003  
LOCUS Sequence 386 from Patent WO02077272.  
DEFINITION AX599046  
ACCESSION AX599046  
VERSION AX599046.1 GI:28399186  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
1  
Bertin, K., Braun, A., Distler, V., Guetig, D., Howe, A., Mueller, J.,  
Olek, A., Piepenbrock, C., Adorjan, P., Grabs, G., Lesche, R., Leu, E.,  
Lewin, A., Lipscher, E., Maier, S., Model, F., Mueller, V., Otto, T.,  
Pellet, C. and Ziebarth, H.  
TITLE  
Methods and nucleic acids for the analysis of hematopoietic cell  
proliferative disorders  
JOURNAL  
Epigenomics AG (DE)  
FEATURES  
source  
1. 8056  
/organism="synthetic construct"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)"  
BASE COUNT 3711 a 0 c 371 g 3974 t  
ORIGIN

Query Match 8.7%; Score 54.4; DB 6; Length 8056;  
Best Local Similarity 46.7%; Pred. No. 1.4;  
Matches 172; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

Qy 162 TTTTGAACAATTTCTACTATATTTTACAGAAATCTTGGTTGAGTTACAGACA 221

Db 6353 TTATATAATAATTTTATTAATTAATTAATAAATAATTAATTAATTAATTAAT 6412

Qy 222 GTCAATATATGCTTTATTAATTAATTAATGCTCAGAGATAGCTATTTATGATGCAAT 281

Db 6413 ATAAATTAATTAATTAATTAATTTTATTAATTAATTAATTTTATTAATTAATTAAT 6472

Qy 282 ATTAATTTTATGCTGTTACGCGATGATTTCAATGCTGACAAAGTTGGCATTAAG 341

Db 6473 AATAAATATGTTTTTTTATTTATTTTAAATTTTGAATAATTTTGAATGAAAA 6532

Qy 342 GACTCAAAAGTTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 401

Db 6533 TGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6592

Qy 402 AATGTTTGTCTCAACAACCTTGTGACTTTTACGTTAGCTTATGTTCAAGTAATGT 461

Db 6593 ATTAATTAATTTTATTAATTAATTAATTAATTTATTTATTTATTTATTTATTAATTAAT 6652

Qy 462 TATAGTATTTGTTCAATATATGATTAATAAAGAGAGTTTCAACTTTTACCATTA 521

Db 6653 TTTTAAATTTTATTAATTAATTAATTAATTAATTTTATTTATTTATTTATTTATTTAT 6712

Qy 522 TTAATGCA 529

Db 6713 ATTATATA 6720

RESULT 13  
PFA929351/c 347050 bp DNA linear INV 29-JAN-2003  
LOCUS PFA929351/c  
DEFINITION Plasmodium falciparum strain 3D7, chromosome 5, segment 1/4.  
ACCESSION AL929351 AL944504  
VERSION AL929351.1 GI:23504490  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
1  
Hall, N., Pain, A., Berriman, M., Churcher, C., Harris, B., Harris, D.,  
Mungall, K., Bowman, S., Atkin, R., Baker, S., Barron, A., Brooks, K.,  
Buckee, C.O., Burrows, C., Cherevach, I., Chillingworth, C.,  
Chillingworth, T., Christodoulou, Z., Clark, L., Clark, R., Cotton, C.,  
Cronin, A., Davies, R., Davis, P., Dear, P., Dearden, F., Doggett, J.,  
Feltwell, T., Goble, A., Goodhead, I., Gwilliam, R., Hamlin, N.,  
Hance, Z., Harper, D., Hauser, H., Hornsby, T., Holroyd, S.,  
Horrocks, P., Humphray, S., Jagels, K., James, K.D., Johnson, D.,  
Kerhornou, A., Knights, A., Konfortov, B., Kyes, S., Larke, N.,  
Lawson, D., Lennard, N., Line, A., Maddison, M., McLean, J., Mooney, P.,  
Moule, S., Murphy, L., Oliver, K., Ormond, D., Price, C., Quail, M.A.,  
Rabinowitz, S., Rajandream, M.A., Rutter, S., Rutherford, K.M.,  
Sanders, M., Simmonds, M., Seeger, K., Sharp, S., Smith, R., Squares, R.,  
Squares, S., Stevens, K., Taylor, K., Tivey, A., Unwin, L.,  
Whitehead, S., Woodward, J., Sulston, J.E., Craig, A., Newbold, C. and  
Barrell, B.G.  
TITLE  
Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13  
JOURNAL  
Nature 419 (6906), 527-531 (2002)  
MEDLINE  
22255708  
PUBMED  
12368867  
REFERENCE  
AUTHORS  
2 (bases 1 to 347050)  
Devlin, K., Baker, S., Davies, P., Mungall, K., Berriman, M., Pain, A.,  
Hall, N., Bowman, S., Churcher, C., Quail, M. and Barrell, B.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (20-SEP-2002) P. falciparum Genome Sequencing Consortium,  
The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus,  
Hinxton, Cambridge CB10 1SA, UK  
COMMENT  
For more information about this sequence or the Malaria Project,  
see [http://www.sanger.ac.uk/Projects/P\\_falciparum](http://www.sanger.ac.uk/Projects/P_falciparum).  
FEATURES  
source  
1. 347050  
/organism="Plasmodium falciparum 3D7"  
/mol\_type="genomic DNA"  
/isolate="3D7"  
/db\_xref="taxon:36329"  
/chromosome="5"  
1. 538  
/note="Telomere"  
11583. 18389  
/note="rep20"  
join(20929. 26310, 27215. 28456)  
/gene="VAR"  
/note="synonym: PPE00054"  
join(20929. 26310, 27215. 28456)  
/gene="VAR"  
/note="Similar to Plasmodium falciparum variant-specific  
surface protein var-2 TR:Q26033 (EMBL:J140609) (2664 aa)  
fasta scores: E(): 2.1e-94, 38.0664 id in 2782 aa"  
/codon\_start=1  
/product="erythrocyte membrane protein 1 (pfeMP1)"  
/protein\_id="CAD51367.1"  
/db\_xref="GI:23504491"  
/translation="MGPGITGTCGCTAGTAKHMFDRIGKOVETVKNVKAENYISELEGL  
SOATLGERVSLKTCQGVEDYRKANGDVARYCANSPVRFSDSESSOCYVNIKD  
NETDNDACGACAPYRRHLCPYNIENKMKSTSTTGHDILAIVOMAKYGDGSKTHYTI  
HKHINDDSAABLCTELASPADIDGILIGKOLYIGDITKKQNGKTEBEKBEINIKRI  
FGKHIVDTNGKKEVLKTRKYDINDPEFFKRLREWMNRAVTAICVHGESDKYF  
RNTGDSHSGTFSQPKYCRGNDGKPEDGANVDPYTPYVPOYLWPFEMWADFC  
RIRHKLMNAKEGCRKXKSGTDRYCSNGYDCTQITGRNLIIVSDSECTGSVCTP



gene  
2668. 3492  
/gene="atpB"  
/locus\_tag="bbp002"  
2668. 3492  
/gene="atpB"  
/locus\_tag="bbp002"  
/EC\_number="3.6.3.14"  
/note="COG0356; protein 6"  
/codon\_start=1  
/transl\_table=1  
/product="ATP synthase A chain"  
/protein\_id="AA026746.1"  
/db\_xref="GI:27903912"  
/translation="MNLHNPNLPNYINHHLLHMLQNLISNLKPLNFNDNYIRNFWCNC  
FDSFISLPIGLVILISFEFKSKTFPIQTENKIQICIELIIDFINNVKKEIYHKNKXLI  
IAPSLTPIRISIKIKIGKGLFDLPFPNPIFPVNPFLIESILSKSVSGSLRFGNM  
FLITFYSIKIKIGKGLFDLPFPNPIFPVNPFLIESILSKSVSGSLRFGNM  
YAGMIFILISGLIPMWLQWILSVPAIFILILISQSFIFWLVILYLSMASTKH"  
3545. 3784  
/gene="atpB"  
/locus\_tag="bbp003"  
3545. 3784  
/gene="atpB"  
/locus\_tag="bbp003"  
/EC\_number="3.6.3.14"  
/note="COG0636; lipid-binding protein;  
dicyclohexylcarbodiimide-binding protein"  
/codon\_start=1  
/transl\_table=1  
/product="ATP synthase C chain"  
/protein\_id="AA026747.1"  
/db\_xref="GI:27903913"  
/translation="MDNVSMDLTYIAVAMIGLAIAGAAGVIGILGSKFLEGVARQPF  
LTSILRQPFVGMGLVDAIPRIAVGLGYMLFAYI"  
3909. 4379  
/gene="atpF"  
/locus\_tag="bbp004"  
3909. 4379  
/gene="atpF"  
/locus\_tag="bbp004"  
/EC\_number="3.6.3.14"  
/note="COG0711"  
/codon\_start=1  
/transl\_table=1  
/product="ATP synthase B chain"  
/protein\_id="AA026748.1"  
/db\_xref="GI:27903914"  
/translation="MDPFWITVGAISFVLFVFECKMYWPSVIFIIETROKEIKDSI  
TFINSKKEILNIFKENSKEIKIIRKNASKIIDSAIOQTOILKOAYLAAREKQITII  
KOALDWIETVOKARVELRQKVSKEIAVEIAKKIINRSICIBEONSIIISLIKKI"  
4392. 4928  
/gene="atpH"  
/locus\_tag="bbp005"  
4392. 4928  
/gene="atpH"  
/locus\_tag="bbp005"  
/EC\_number="3.6.3.14"  
/note="COG0712"  
/codon\_start=1  
/transl\_table=1  
/product="ATP synthase delta chain"  
/protein\_id="AA026749.1"  
/db\_xref="GI:27903915"  
/translation="WSSTKDELQLYAKATINCAISNHSQSLDHWKTMLQMANIINNE  
IKNLSIKVAFQHVAISFLIDCCNKRVQYGNLKIILAEENRMLLEKLYAFKIVLC  
LYQGVNTVYSAHLNBEYISKINIMLKKRFKKINVTYIDESIIIGGLIIFKCDTV  
INASHLSLEKILNIIQY"  
4942. 6486  
/gene="atpA"  
/gene="atpA"

CDS /locus\_tag="bhp006"  
 4942..6486  
 /gene="atpA"  
 /locus\_tag="bhp006"  
 /EC\_number="3.6.3.14"  
 /note="COG0056; FOF1-type ATP synthase alpha subunit"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ATP synthase subunit alpha"  
 /protein\_id="AA026750.1"  
 /db\_xref="GI:27903916"  
 /translation="MQISSSEICELISKIKKPIISSMNEGVISVSDGIITVGL  
 SDVQGMELSLPSDKVAIALNLERNVGAIVMGSEYTHITGETKLIISTGRLEIVSGK  
 FLGRVINALGPIDGKSIQIEKFLPVEINAPGVIDRQKISEPLQTGKSIDAMVPIG  
 KGRELLIGDROGKTSIAIDTIINORTKIKCIYVAGKSTIVNLVROLENOAL  
 NHTIVIVASASEAALQVLYVSGSGISGEFPRDQKALIVDDLKHAIVROISLL  
 LRPRGEARPRGDI EYLAHRLERACRVNSMYLKNITGNKSTGSLTALPIETED  
 GDVSSFTPTNVISTIDGQIFLESNLVNSGIPALNPGISVRVGAACQITRLSG  
 IRTLAQYQELIARFOSSELDERTNQLHGKLIILKQKQYHPMSIAQAILERA  
 AENNRLDVSEQIIEKEKMLLFENTNSELVLSINNYKINDVVENQLSDVINAEK  
 LTKCMS"  
 gene 6524..7399  
 /gene="atpC"  
 /locus\_tag="bhp007"  
 6524..7399  
 /gene="atpC"  
 /locus\_tag="bhp007"  
 /EC\_number="3.6.3.14"  
 /note="COG0224"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ATP synthase gamma chain"  
 /protein\_id="AA026751.1"  
 /db\_xref="GI:27903917"  
 /translation="MIGRIEIRSKCKSINNTKITYKAMEMVSIKAKIKRMSSSRP  
 YPAINOVISHVITGNLBHHTYFNQKVKIGIVISTDGLCGNLMTLFFKVLVY  
 LTERINHILNLLNFVIGKALTFKFSFTNNVFEISLKNDFKIDIMEMIRISLEMY  
 ISGKIDKFLAVNKNSTIIQTPTLVOLPLIKPKLGKKEVKTWDYIYESNSKLLN  
 VLNRYIEFOIYOSILENLVCEASRLMAMKQATDNSADLLKALQNNVKNYKQSSIIQ  
 ELTEIISGAAVSLN"  
 gene 7427..8824  
 /gene="atpD"  
 /locus\_tag="bhp008"  
 7427..8824  
 /gene="atpD"  
 /locus\_tag="bhp008"  
 /EC\_number="3.6.3.14"  
 /note="COG0035"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ATP synthase beta chain"  
 /protein\_id="AA026752.1"  
 /db\_xref="GI:27903918"  
 /translation="WITGKIVQIIGAVVDFEQQSVPKIFNALKVDNQSILILEVO  
 QQLSGIVRTIAMGSSNGLKRGGLVLDLEHGIKVPVGTALGRIVNLGQPIDKQGL  
 KMHNSDI EYWEIHRKAPSEQLTSEYLETGKVIDLIDLPFGKGAQVFGGAGG  
 KTVMMELIRNLITTEHSGSYFTGVBRETRGNDPYHMSDRVLDRVSLVYGGMNP  
 PGNRLVAFGLITAEKPRNGCHVLLPIMIVRYTLAGTYSVALGRISBAVQRT  
 LSEMGVLOERITTKGTSITSAIVYPADLDLDPSPATTFSHLDSTITLSROIVL  
 GIYPAIDPLNSTKQDPRIVGQHYVALGVRSLIRYQDLKQIILAILGDELSELSD  
 KIIIVSRARKIQKPLSQPFVAEIFTGSGKVKVQLQDTINGFKDIIIEGKVHVPQARY  
 MVSINVEIKSKL"  
 gene 8826..9263  
 /gene="atpC"  
 /locus\_tag="bhp009"  
 8826..9263  
 /gene="atpC"  
 /locus\_tag="bhp009"  
 /EC\_number="3.6.3.14"  
 /note="COG0355"  
 /codon\_start=1  
 /transl\_table=11

Query Match 8.7%; Score 54.2; DB 1; Length 310003;  
 Best Local Similarity 46.4%; Pred. No. 0.53;  
 Matches 213; Conservative 0; Mismatches 243; Indels 3; Gaps 1;

Qy	163	TTTGAACATTTTCTACATATTTTACAGATCATTTGCGTTGAGGTTACGACACG 222
Db	261562	TTTAAAGATTTTCTATTAAGTTTACTTTTAATATCTATTTTCAAAAGATTTTA 261621
Qy	223	TCAGTATATGCTTATTTATTTATGTTGCTCAAGAGATAGCTATTTATGATGCAATA 282
Db	261622	TTTTAGATCAGATTTCTTATATATTTTTCACATATCTTATTTCTATTTAGANATA 261681
Qy	283	TTACTTTTATTTGCTGTAGCGCATGATTCATTAAGCTGTACAAAGTTGTCATTAAGC 342
Db	261682	ATTTTTCCTATAGTTAATTTTATAGTTTCCGCAATGATATATAGTTGTTCAATTTT 261741
Qy	343	ACTCAAAAGTTTAAATTTAAATGAGATTAATTTATATATTAATAAAGGTTGAAGGA 402
Db	261742	AAITTTAAATTTAATAAATTTGATATTAATAATAATCAATATTAAGATTTAAGAAAT 261801
Qy	403	ATGTTGCTTCTCAACAAACCTTGTGACTTTTACGTAGTTAGTTCAAGTATAGTT 462
Db	261802	GTTTTCCTATTAAGATATTTGATTTATTTTAAACAGCTTAA--AATAGTTTT 261858
Qy	463	ATAGATATGTTCCATATATGATTAATAAAGAGATTTTCAACTTTTACATTAATAT 522
Db	261859	GTCGATATTTAGATATATTTATTTATATCTAATAATTTTAAAGATATGCTTTCCCA 261918
Qy	523	TATCAAGTCTTCAAGTGTGCGCATTAATATGCTTAATACAGAAATGATCTGTTTAA 582
Db	261919	AAITTTCAATTTGATTTAATAATTAATAATTAATGATGACACTATTTTGTGTTTTTGG 261978
Qy	583	TATACGTAATTCCTATGACATATATGACGTGACAT 621
Db	261979	ACTATTTAATATTTTATTTTATTCATATTAATATGAT 262017

**RESULT 15**  
 PFA929351 347050 bp DNA linear INV 29-JAN-2003  
 LOCUS Plasmodium falciiparum strain 3D7, chromosome 5, segment 1/4.  
 ACCESSION AL929351 AL644504  
 VERSION AL929351.1 GI:23504490  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Plasmodium falciiparum 3D7  
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

**REFERENCE**  
 1  
 Hall, N., Pain, A., Berriman, M., Churcher, C., Harris, B., Harris, D.,  
 Mungall, K., Bowman, S., Ackin, R., Baker, S., Barton, A., Brooks, K.,  
 Buckee, C. O., Burrows, C., Cherevach, I., Chillingworth, C.,  
 Chillingworth, T., Christodoulou, Z., Clark, L., Clark, R., Cotton, C.,  
 Cronin, A., Davies, R., Davis, P., Dear, P., Dearden, F., Doggett, J.,  
 Fellwell, T., Goble, A., Goodhead, I., Gwilliam, R., Hamlin, N.,  
 Hance, Z., Harper, D., Hauser, H., Hornsbey, T., Holroyd, S.,  
 Horrocks, P., Humphray, S., Jagels, K., James, K. D., Johnson, D.,  
 Keshornou, A., Knights, A., Konfortov, B., Kyes, S., Latke, N.,  
 Lawson, D., Lemard, N., Line, A., Maddison, M., McLean, J., Mooney, P.,  
 Moule, S., Murphy, L., Oliver, K., Ormond, D., Price, C., Quail, M. A.,  
 Rabbittowitsch, E., Rajandream, M. A., Rutter, S., Rutherford, K. M.,  
 Sanders, M., Simmonds, M., Seeger, K., Sharp, S., Smith, R., Squares, R.,  
 Squares, S., Stevens, K., Taylor, K., Tivey, A., Unwin, L.,  
 Whitehead, S., Woodward, J., Sibleon, J. E., Craig, A., Newbold, C. and  
 Barrell, B. G.  
 Sequence of Plasmodium falciiparum chromosomes 1, 3-9 and 13  
 JOURNAL Nature 419 (6906), 527-531 (2002)  
 MEDLINE 22255708  
 PUBMED 12368867  
 2 (bases 1 to 347050)  
**AUTHORS**  
 Devlin, K., Baker, S., Davies, P., Mungall, K., Berriman, M., Pain, A.,  
 Hall, N., Bowman, S., Churcher, C., Quail, M. and Barrell, B.  
**TITLE**  
 Direct Submission  
**JOURNAL**  
 Submitted (20-SEP-2002) P.falciiparum Genome Sequencing Consortium,



```

gene
/codon_start=1
/complement (29923..30963)
/gene="RIF_degenerate"
/note="synonym: PFE0015c"
/pseudo

```

CDS  
 complement (29923..30963)  
 /gene="RIF, degenerate"  
 /note="Similar to Plasmodium falciparum rifin pfbl040w  
 TR:096293 (EMBL:AE001433) (345 aa) fasta scores: E():  
 1.1e-50, 53.161% id in 348 aa. Contains nonsense mutations  
 at apparent start position for exon 2"  
 /pseudo  
 /codon\_start=1  
 complement (30904..30909)  
 /gene="RIF, degenerate"  
 /note="Possible splice donor"  
 /pseudo  
 complement (join(33056..34126,34325..34378))  
 /gene="RIF"  
 /note="synonym: PFE0020c"  
 complement (join(33056..34126,34325..34378))  
 /gene="RIF"  
 /note="Similar to Plasmodium falciparum rifin pfB0030C  
 TR:096112 (EMBL:AE001367) (370 aa) fasta scores: E():  
 3.3e-62, 53.125% id in 384 aa  
 1 probable transmembrane helix predicted for by TMHMM2.0  
 at aa 332-354"  
 /codon\_start=1  
 /product="RIFIN"  
 /protein\_id="CAD51370.1"  
 /db\_xref="GI:23504492"  
 /translation="MKPSYENILLFSTPLNLIINDHSKYSCKHTSNKTKTPHRSILY  
 ECGISPPANNDDPEMKRYWOCPEFDRIORPHEVDENRMSKMOCKECDEIKIILIL  
 KDGLKQKQEQOLNLTETKIITDDIPTCYCEKSLADKYKGLRCGYGGLTAPVGLL  
 GAVVAVHWKQPALAEALIALAKSAEATISALANAAGIOAGKIAYIBSKIKLYVYFW  
 EMSVYILMSHYNGVNIETALIEHPKFEVCKDAEVLIDCKNADPMGGLIKGSGVTT  
 GLPLPDAPRVKIGVQAGEAPVADAAQOTVAETLEKETAINTIFMSKOTAILI  
 ASVAIVIVITLIIYILIRYRRKKKKKLLQYIKKE"  
 complement (join(36464..37324,37496..37576))  
 /gene="RIF"  
 /note="synonym: PFE0025c"  
 complement (join(36464..37324,37496..37576))  
 /gene="RIF"  
 /note="Similar to Plasmodium falciparum rifin pfbl005W  
 TR:096288 (EMBL:AE001432) (316 aa) fasta scores: E():  
 8.3e-52, 54.574% id in 317 aa  
 2 probable transmembrane helices predicted for by TMHMM2.0  
 at aa 122-144 and 271-293"  
 /codon\_start=1  
 /product="RIFIN"  
 /protein\_id="CAD51371.1"  
 /db\_xref="GI:23504493"  
 /translation="MAINMKLTKSKILPLSLQNLITSSYVNNKPKYTPHTATL  
 SRVNLSECDLNMVNDKADMKSVKSNPEPORSREKYEEMRTTRQYKCECDI  
 KILLDKIEKLIEKVEKCCILCGGJAGGVAPVGLSGGLANEMKKAAYAAVDV  
 IKAIAKEGSLPYIADGSTMETWTKINIGNTSQKSLVEITIIILNKKEEDALAGP  
 PKAEAEATSEGEVEFGSINSRMANAAEAAGNANGKTAEMTSVGTICSNPVVIS  
 IYVIVIAVILIIYILIRYRRKKKKKQYIKLNLQ"  
 complement (39443..40488)  
 /gene="STEVOR, degenerate"  
 /note="synonym: PFE0030c"  
 /pseudo  
 complement (39443..40488)  
 /gene="STEVOR, degenerate"  
 /note="Apparent single frameshift in exon II"  
 /pseudo  
 /codon\_start=1  
 complement (41129..41403)  
 /gene="RIF, degenerate"  
 /note="synonym: PFE0035c"  
 /pseudo  
 complement (41129..41403)



[illegible]

JOURNAL. Nature 419 (6906), 498-511 (2002)  
PUBMED. 12368864  
REFERENCE 2 (bases 1 to. 257757)  
AUTHORS Gardner, M.J.  
TITLE Direct Submission  
JOURNAL Submitted (13-SEP-2002) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA  
FEATURES  
source  
1. 257757  
/organism="Plasmodium falciparum 3D7"  
/mol\_type="genomic DNA"  
/isolate="3D7"  
/db\_xref="taxon:36329"  
/chromosome="11"  
257. .381  
repeat\_region  
/rpt\_type=tandem  
305. .373  
repeat\_region  
/rpt\_type=tandem  
763. .787  
repeat\_region  
/rpt\_type=tandem  
1304. .1333  
repeat\_region  
/rpt\_type=tandem  
complement (<1434. .>1994)  
/locus\_tag="PF11\_0070"  
complement (join(<1434. .>1994))  
/locus\_tag="PF11\_0070"  
complement (join(1434. .1828,1910. .1994))  
/locus\_tag="PF11\_0070"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="PAAJ35659.1"  
/db\_xref="GI:23495997"  
translation="MCMYVCIIETEKKKKKKKKKLPTIHKIDVNNKAEONPL  
VVRISGKHLFEKROPSFLYPTVITILMMRDYTKLHAVAVAYERKSLFPYLGIKK  
ANVIYINKKPKRIWENIKSIFSITAVTNSMDEFTPTKTLFLFISLSTKIV"  
1872. .1902  
repeat\_region  
/rpt\_type=tandem  
1926. .1960  
repeat\_region  
/rpt\_type=tandem  
2199. .2240  
repeat\_region  
/rpt\_type=tandem  
2384. .2542  
repeat\_region  
/rpt\_type=tandem  
2582. .2650  
repeat\_region  
/rpt\_type=tandem  
complement (<2647. .>4074)  
/locus\_tag="PF11\_0071"  
complement (<2647. .>4074)  
/locus\_tag="PF11\_0071"  
complement (2647. .4074)  
/locus\_tag="PF11\_0071"  
/codon\_start=1  
/product="RuvB DNA helicase, putative"  
/protein\_id="PAAJ35660.1"  
/db\_xref="GI:23495998"  
translation="MEINISINITSSNNKKERVNIHSIKLGCVNTIYLTENEISIS  
DEKSYMPFNSCGLVGQFARBAEALFLVLDIIROKLAGCIILAGSGSGKSLAIGI  
SRRLNKKPVPFLSGEYVSNELKTEVILEARSKIHIIKKEKLVSEGEVVDVVE  
ENGCISLNKAKOINAIITLKSVKSKTILRLPKHEOJARAKITIGDIYIETNG  
HVRKLGCDNYAKEDIDEPBYVSLPGEVHKKEVVOQSLHDIIDANANPVEGBL  
ASVNLGSLRKPKETIEKRLVEIKTVNKFLESGLAEIIPGVYIDEANHLDECEYS  
LNAISFLPAPILIMATNGRICVKGIDNIEPGIIEVDLIDRLIIIKTEPYTLKEIVQ  
ILALRAOTEKINITEGMYLAKIGOSSIRFAMLLPESRIATLETGNTIIDIKHEI  
QAELEPDAKTSNHRVVDGSKFVN"  
4159. .4258  
repeat\_region  
/rpt\_type=tandem  
4159. .4215  
repeat\_region  
/rpt\_type=tandem  
4718. .4749  
repeat\_region  
/rpt\_type=tandem  
4922. .4978  
repeat\_region  
/rpt\_type=tandem  
5175. .5199  
repeat\_region

```
repeat_region      /rpt_type=tandem
                    5322..5352
                    /rpt_type=tandem
repeat_region      5350..5397
                    /rpt_type=tandem
repeat_region      5405..5438
                    /rpt_type=tandem
repeat_region      5531..5557
                    /rpt_type=tandem
gene               /rpt_type=tandem
                    <5712..>6914
mrna               /locus_tag="PF11_0072"
                    join(<5712..5802,5896..6191,6311..6454,6636..>6914)
                    /locus_tag="PF11_0072"
                    join(5712..5802,5896..6191,6311..6454,6636..6914)
CDS                /locus_tag="PF11_0072"
                    /codon_start=1
                    /product="hypothetical protein"
                    /protein_id="AAN35661.1"
                    /db_xref="GI:23495999"
                    /translation="MKNTKSCVTCILFPYIOIWINCLFCFGYNINKNLGVYPSLIRK
                    KNPPEKRRQILIKQNGKENIFPDQNTNTIKKDKTKQTKTNDIKDKIKKEKIEI
                    NLKKSSENILERSFSFKEDITIKPELKEKLIQTKRKIFORHDKCGSSEIOIIL
                    LTFKIFPLEHKKNKDPACRGLFKCVSKRRRLVYLGRDREMEFKTISYENIKK
                    PLRPTPEYVNDLKYIHFNNTKRFKNNAEKKKDKLKKKQVTDKILFNG"
```

```
repeat_region      6250..6288
                    /rpt_type=tandem
repeat_region      6482..6530
                    /rpt_type=tandem
repeat_region      6484..6533
                    /rpt_type=tandem
repeat_region      6913..6938
                    /rpt_type=tandem
repeat_region      7132..7156
                    /rpt_type=tandem
repeat_region      7287..7327
                    /rpt_type=tandem
repeat_region      7691..7726
                    /rpt_type=tandem
gene               /rpt_type=tandem
                    <7806..>10244
mrna               /locus_tag="PF11_0073"
                    join(<7806..9712,9978..10049,10155..10162,10204..>10244)
                    /locus_tag="PF11_0073"
                    join(7806..9712,9978..10049,10155..10162,10204..10244)
CDS                /locus_tag="PF11_0073"
                    /codon_start=1
                    /product="hypothetical protein"
                    /protein_id="AAN35662.1"
                    /db_xref="GI:23496000"
                    /translation="MFCSPFLFFFFVONKEIALILCSKVLKKGHIYPRKSHFHI
                    YSVNNVNTLKKHKNYSKLDKRIKIEGVHNNKKKIYELDLIKNNNDVYIEDH
                    KDKNIDNNNNKKDKDDINNDKKNKINDTPAONTINDKRYSNPEYIENNLLK
                    LKNLLEVSSTYINKSDVILVPLNHDILKHLIELIPENNLIINSLDYIFLEKNNYK
                    SDININNNNNNNNNKYNICPPYNNIYNIIFPIYVNDQDPLMISSTPSEFTNKK
                    DNIHKEKQOLVHVIQKFLNDHYNLKDQEHKKYININFLFTKNDILDKDKDIEE
                    NMLYYPKKNIILKEKNTVNTNTNTNTKQDONLILKVHKNNDTFPEKKYVPTLKN
                    LTNENYVFINPLSKYKDLIKEXIYIADITIDPNNNVKHIQININIKKKIHIELLN
                    NFKKCKLNKINSHNVGRIKKQYISYLHITLSLDEKIDENELIKKYLQVKEKN
                    IQKIKINSDILQFKKQIINHDLQFDYNSKFIPLSNKNVQAQGSYITTYIKL
                    LQILNQIANEIIINYISKGLVKNYSFNEFSKRYTLINVTITTYIEMLNKKNVINE
                    NLSNPASFSSYKDDISLSPKKDIMITTSEMDQVQKIYIQISLVVPTPKRCS"
```

```
repeat_region      8865..8890
                    /rpt_type=tandem
repeat_region      9936..9975
                    /rpt_type=tandem
repeat_region      10060..10092
                    /rpt_type=tandem
repeat_region      10372..10409
                    /rpt_type=tandem
repeat_region      10454..10485
                    /rpt_type=tandem
gene               /rpt_type=tandem
                    complement(<10677..>13503)
                    /locus_tag="PF11_0074"
                    complement(join(<10677..10818,10922..10993,11120..11199,
```

```
11295..11397,11559..12736,12847..>13503))
                    /locus_tag="PF11_0074"
                    complement(join(10677..10818,10922..10993,11120..11199,
```

```
11295..11397,11559..12736,12847..13503))
                    /locus_tag="PF11_0074"
                    /codon_start=1
                    /product="hypothetical protein"
                    /protein_id="AAN35663.1"
                    /db_xref="GI:23496001"
                    /translation="MNPFLSNVLCITNNMAYCISLKKYKRSYNLKNRKHQONV
                    YGIPRMKWLISYIPVIREELINNEKKSVDIFIDNGNTHHTHANKELPTIDHE
                    ELFENIIQYVKNLFFYLKPKKLIVIGDVSPPAKNQORRRRLSLFKINDNNTSN
                    LENENCITTTGDEFWYKINLSINKFKILKKKXVEPDVIFSGDVAAGEKRLIKYLR
                    ENCKRDNPKYNNHCICYGLADJLMLSLVTHNNI FILRDKFKISNDLHNMLENIE
                    TONFNNNDVNSNEPONDNIINVEBEINOLNEKSKTNSLVEKNVEYENNEIN
                    NEINNEINNEINNEINNEINNEINNEINNEINNEINNEINNEINNEINNEINNE
                    INNEINNEINNEINNEINNEINNEINNEINNEINNEINNEINNEINNEINNEINNE
                    NSIKTQIATYINKLREKKNIVFISISRVDDIVFSLVAGNDLPHIIPRIDINSGMNE
                    ILSNVIPIYKYSNVIYIKOKVHIERLILKILASQEFYFKRGINENISEPTDEQ
                    KYKKYVILHKEGLEDPEIKQIVKRYIEGLFWMNLHYHFGCASWMEYPIYHVALCD
                    LSPKSDPFEPEKYYKKELFMGKPYSAFTHLISVLPQDKNLDPAYKNIVDEVK
                    SFPENYKIDENKGETWEYIVHLPIFNCMNINKIILEKSTIKLAKLYNCKEQ
                    IYL"
```

```
repeat_region      11438..11479
                    /rpt_type=tandem
repeat_region      12185..12400
                    /rpt_type=tandem
repeat_region      12784..12816
                    /rpt_type=tandem
repeat_region      13620..13652
                    /rpt_type=tandem
repeat_region      13757..13788
                    /rpt_type=tandem
```

```
Query Match      8.6%; Score 53.8; DB 3; Length 257757;
Best Local Similarity 47.9%; Pred. No. 0.66; 232; Indels 3; Gaps 2;
Matches 216; Conservative 0; Mismatches 232; Indels 3; Gaps 2;
```

```
QY      135 TATTATTACAGCGGTAATGGACGTCATTTGGAACAATTTTCTACATATTTTATACGA 194
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      108071 TATTAAGTTAATGCGAATAAATATCTTGATATAAATATATATATATATANA 108130

QY      195 ATCAATTCGGTTGAGTTACAGACAGTCAGTATATGCTTATTTATTTATGTTGCTCA 254
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      108131 AGGATATCCTTAGATGAGACAAATAATATATATATATATATATATATAT 108190

QY      255 AGAGATGACTATTTATGATGCCAATTTACTTTTATGCTGTACAGCGATGCAATTC 314
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      108191 ATATATATTTTATTTATTTATATATATATATATATATATATATATATATG 108250

QY      315 ATTACGTCAGTACAGT--TGGTGCAATTATGACCTCAAGAGTTTAAATTAATGAGAT 372
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      108251 GTTACACATATTTATATATAAACAATATATAAATAAATAAAGACACAGAACTTAAGTCCCA 108310

QY      373 AAATTTATATATAAAGAGGTTGAAAGAAATGTTGCTTCACAAACACTTGTTGCA 432
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      108311 AAAGAAGAAACAATAAACAATCTAATCAATGCGTGTGTATATCAAGAGTTAAAGGA 108370

QY      433 CTTTACGTAGTTAGTTACCAATTTGTTATAGGATTTGTCCTATATATGATTTAAAA 492
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      108371 GCTATATTTAAGTGAAGAAATAATTTGAGAGAAATGAAAAAATTTAGAAATTTAT 108430

QY      493 GGAGAGCTTTCAAACTTTTACCATTAATATATGCAAGTCTTCAGTGTGCGA--GATTA 551
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      108431 AGCGAATAAAGAAAGGCTTAATATGTTTTTTAGGTATTCATATAAATGATACATGGGAA 108490

QY      552 TATGCTTATATACAGGAATAGTACTGTTTAA 582
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      108491 GATGCTTATATATATATAGAAATGTTTAA 108521
```

```
RESULT 17
AXS58988
```

LOCUS AX598988 1501 bp DNA linear PAT 14-FEB-2003  
DEFINITION Sequence 328 from Patent WO02077272.  
ACCESSION AX598988  
VERSION AX598988.1 GI:28399128  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
REFERENCE  
1  
AUTHORS  
1  
Bellin, K., Braun, A., Distler, J., Guetig, D., Howe, A., Mueller, J.,  
Olek, A., Piepenbrock, C., Adorjan, P., Grabs, G., Lesche, R., Leu, E.,  
Lewin, A., Lipscher, E., Maier, S., Model, F., Mueller, V., Otto, T.,  
Petek, C. and Ziebarth, H.  
TITLE  
Methods and nucleic acids for the analysis of hematopoietic cell  
proliferative disorders  
JOURNAL  
Patent: WO 02077272-A 328 03-OCT-2002;  
Epigenomics AG (DE)  
FEATURES  
source  
1. .1501  
/organism="synthetic construct"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)"  
BASE COUNT 604 a 0 c 130 g 767 t  
ORIGIN  
Query Match 8.6%; Score 53.6; DB 6; Length 1501;  
Best Local Similarity 47.7%; Pred. No. 3;  
Matches 187; Conservative 0; Mismatches 204; Indels 1; Gaps 1;  
QY 131 TTGCTATTATTCAGCGGTAATGAGCGTCAATTTGAAACAATTTCTACTATATTTTA 190  
DB 669 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 728  
QY 191 CAGATCATTTGCGTTGAGGTACAGACAGCATATGCTTTATTTATTTATGTTG 250  
DB 729 TTGTTTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 788  
QY 251 CTCAGAGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 310  
DB 789 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 848  
QY 311 TTTCATTAAGTGAAGGTTGTCATTATGACTACAAAGGTTTAAATTAATTAATTAATTA 370  
DB 849 TTTAAATATGTTGAATGTTGTAATAAATTAATAAATTAATAATTAATTAATTAATTA 908  
QY 371 GTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 430  
DB 909 TTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 967  
QY 431 GACTTTAAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 490  
DB 968 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1027  
QY 491 AAGGAGGTTTCAACTTTTACCATTAAT 522  
DB 1028 AAATGTAATTTTAAATTTAATTTAATTT 1059  
RESULT 18  
BD092427/c 945 bp DNA linear PAT 27-AUG-2002  
LOCUS BD092427/c  
DEFINITION Identification of polynucleotides encoding novel helicobacter  
polypeptides in the helicobacter genome.  
ACCESSION BD092427.1 GI:22638038  
VERSION BD092427.1  
KEYWORDS JP 2001527393-A/208.  
SOURCE synthetic construct  
ORGANISM synthetic construct  
REFERENCE  
1 (bases 1 to 945)  
AUTHORS Kleinhous, H., Garawi, A.A., Miller, C., Tomb, J.F. and Oomen, R.P.  
TITLE Identification of polynucleotides encoding novel helicobacter

JOURNAL  
polypeptides in the helicobacter genome  
Patent: JP 2001527393-A 208 25-DEC-2001;  
MERIEUX ORAVAX SOCIETE EN NOM COLLECTIF PASTEUR MERIEUX SERUMS ET  
VACCINS AGROBIOLOGICAL RESOURCES MINISTREY O SA, HUMAN GENOME  
SCIENCES INC  
COMMENT  
PN JP 2001527393-A/208  
PD 25-DEC-2001  
PF 01-APR-1998 JP 1998541947  
PR 01-APR-1997 US 08/833457,24-JUN-1997 US 08/881227 PR  
29-JUL-1997 US 08/502615  
PI HAROLD KLEINHOUSS, AMAL AL GARAWI, CHARLES MILLER, JEAN FRANCOIS  
PI TOMB,  
PI RAYMOND PETER OOMEN  
PC A01N43/04,A61K31/70  
CC Strandedness: Single;  
CC Topology: Linear;  
FH key  
FEATURES  
source  
1. .945  
/organism="synthetic construct"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32630"  
BASE COUNT 298 a 186 c 169 g 292 t  
ORIGIN  
Query Match 8.5%; Score 53; DB 6; Length 945;  
Best Local Similarity 46.9%; Pred. No. 4.4;  
Matches 201; Conservative 0; Mismatches 225; Indels 3; Gaps 1;  
QY 19 AAAACAGAAAGCAACCCGAAAGCAGCGTCAAGAGCTGTTGAAGGAGTGTCCT 78  
DB 845 AAAACCAAGCTCCCTACCGGAAATCCAAAGCCAGAGAAAGCAATGTCCT 786  
QY 79 AAATCAGAGAGGTACTTAAGCATGACTGACTGACAGAGGATGCGGCTGCTATT 138  
DB 785 AAGACATGAGAGTGTGGGGGTTTGGGTTATTTGCGGCTAATTAATTTTGT 726  
QY 139 TATTCAGCGTAATGAGACGCTATTTTGAACAATTTTCACTAATTTTACAGAAATCA 198  
DB 725 TTTTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 669  
QY 199 TTTCGTTTGAAGTTACAGACAGCTCATATGCTTATTTATTTATTTATTTATTTATTT 258  
DB 668 TTTTCCCTAGATTTCAATTAAGAAAGGCTTCAAGAGCTGTTTAACTGCTAAAGAC 609  
QY 259 ATAGCTATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 318  
DB 608 ACTTTTATTTGCTTTTACGATTTTATATCAATTTTATGATGATGATGATGATGATG 549  
QY 319 CGTGTACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 378  
DB 548 GTCTTGCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 489  
QY 379 AATTAATTAAGAGGTTGAAGAAAGATTTGCTTCTCAACAAACCTGTTGACCTTTTA 438  
DB 488 AACCCTATCAATGCGTCAAAACCTTTTCTTTAAAGAGCTGTTGATGAGGATTTG 429  
QY 439 CGTATGTTA 447  
DB 428 ATCAGCTTA 420  
RESULT 19  
AE001502 10525 bp DNA linear BCT 20-JAN-1999  
LOCUS AE001502  
DEFINITION Helicobacter pylori, strain J99 section 63 of 132 of the complete  
genome.  
ACCESSION AE001502 AE001439  
VERSION AE001502.1 GI:4155264  
KEYWORDS  
SOURCE Helicobacter pylori J99  
ORGANISM Helicobacter pylori J99  
Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;



[illegible]

QY	379	AAATATATAAAGGGTGAAGCAATCTTGGCTCTCAACAAACACTGTTCACCTTTA	438
Db	1044	AAACCCATCATGAGCGGTCAAAACCTTTTCTTTCTTAAAAAAGATCCTGATGGAGTTTG	1103
QY	439	CGTAGTTTA	447
Db	1104	ATCACCTTTA	1112
RESULT 20			
LOCUS	AE000589		
DEFINITION	Helicobacter pylori 26695 section 67 of 134 of the complete genome.		
ACCESSION	AE000589	AE000511	
VERSION	AE000589.1	GI:2313895	
KEYWORDS			
SOURCE			
ORGANISM	Helicobacter pylori 26695		
REFERENCE	Helicobacter pylori 26695		
AUTHORS	Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales; Helicobacteriaceae; Helicobacter.		
	1 (bases 1 to 10860)		
	Tomb,J.-F., White,O., Kerlavage,A.R., Clayton,R.A., Sutton,G.G.,		
	Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S.,		
	Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F.,		
	Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G.,		
	Glodek,A., McKenney,K., Fitzgerald,L.M., Lee,N., Adams,M.D.,		
	Hickey,E.K., Berg,D.E., Gocayne,J.D., Uiterback,T.R.,		
	Hickey,E.K., Berg,D.E., Gocayne,J.D., Uiterback,T.R.,		
	Peterson,J.D., Kelley,J.M., Karp,P.D., Smith,H.O., Fraser,C.M. and		
	Venter,J.C.		
TITLE	The complete genome sequence of the gastric pathogen Helicobacter pylori		
JOURNAL	Nature 388 (6642), 539-547 (1997)		
MEDLINE	97394467		
PUBMED	9252185		
REFERENCE	2 (bases 1 to 10860)		
AUTHORS	Tomb,J.-F., White,O., Kerlavage,A.R., Clayton,R.A., Sutton,G.G.,		
	Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S.,		
	Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F.,		
	Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G.,		
	Glodek,A., McKenney,K., Fitzgerald,L.M., Lee,N., Adams,M.D.,		
	Hickey,E.K., Berg,D.E., Gocayne,J.D., Uiterback,T.R.,		
	Peterson,J.D., Kelley,J.M., Cotton,M.D., Weidman,J.M., Fujii,C.,		
	Bowman,C., Wathey,L., Wallin,E., Hayes,W.S., Borodovsky,M.,		
	Karp,P.D., Smith,H.O., Fraser,C.M. and Venter,J.C.		
	Direct Submission		
TITLE	Submitted (06-AUG-1997) The Institute for Genomic Research, 9712		
JOURNAL	Medical Center Dr, Rockville, MD 20850, USA		
	3 (bases 1 to 10860)		
REFERENCE	White,O.		
AUTHORS	Direct Submission		
TITLE	Submitted (17-MAR-1999) The Institute for Genomic Research, 9712		
JOURNAL	Medical Center Dr, Rockville, MD 20850, USA		
	Location/Qualifiers		
FEATURES	1. 10860		
source	/organism="Helicobacter pylori 26695"		
	/mol_type="genomic DNA"		
	/strain="26695"		
	/db_xref="taxon:85962"		
	/complement (73. 147)		
gene	/gene="HP0767"		
CDs	/complement (73. 147)		
	/gene="HP0767"		
	/note="hypothetical protein; identified by Genemark;		
	putative"		
	/codon_start=1		
	/transl_table=11		
	/product="H. pylori predicted coding region HP0767"		
	/protein_id="AAD07827.1"		
	/db_xref="GI:2313906"		
	/translation="MEKVIAPFSPFISFSGGGYL"		
	complement (251. 1216)		
gene	/gene="HP0768"		



Db 1917 AAAACCGACTCCCTAGCCGCGCAAAAAATCCAAAAAGCCAGAGAAAGCAATGCTCT 1976  
Qy 79 AAATCAGAGAGGTTACTTAAGACATGACTCTGCGAGAGGAGTCTGGGCTTGTCTAT 138  
Db 1977 AAGAGCATGAGAGTGGGGGTTTGGGGTTATGGCGGGCTAATAGATTTTTCIT 2036  
Qy 139 TATTCAGGCGTAAAGGACGTCATTTTGAACAAATTTTCTACTATTTTATACAGATCA 198  
Db 2037 TTTTATATATGTTGGGTGAGTGGCTTTAGCGAAATGTATCCCATGTCCTAAGA--T 2093  
Qy 199 TTTGCGTTGAGTTACAGCAGTCACTATATGCTTATTTATTTATTTATGTTGCTCAAG 258  
Db 2094 TTTTCCTAGATTTCACTAAAGAAAGCGTTCAAGGCTGTTTACCACTGCTTAAAGAC 2153  
Qy 259 ATACCTATTTTATGATGCCAATATCTTTTATTTGCTGTACGCGAGATTTTCATTA 318  
Db 2154 ACTTTTATTTGCTTTTACCGATTTTATCATTTTATAGTGGTGGCTTTTATCTAT 2213  
Qy 319 CGGTACAGTTGGTGCATTTATGACCTAACAAGTTTAAATTTAAATGAGTAAATTT 378  
Db 2214 GCTTTCGCAATTTGGCTGGCTCTTGGCCCTTAAAGCATGAGCTTAAATTTTCTPAAATC 2273  
Qy 379 AATATAATAAAGGTTGAAGAGATGTTGCTTCTCAACAACACTTGTTCGACTTTA 438  
Db 2274 AACCTATCAATGCGGTCAAAAACCTTTTCTTTAAAAAGCTCTTGATGGAGATTG 2333  
Qy 439 CGTAGTTTA 447  
Db 2334 ATCACTTA 2342

RESULT 21  
LOCUS AX345303 7167 bp DNA linear PAT 01-FEB-2002  
DEFINITION . Sequence 374 from Patent WO0200928.  
ACCESSION AX345303  
VERSION AX345303.1 GI:18493189  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
REFERENCE 1  
AUTHORS Ojek.A., Piepenbrock.C. and Berlin.K.  
TITLE Diagnosis of diseases associated with the immune system  
JOURNAL Patent: WO 0200928-A 374 03-JAN-2002;  
Epigenomics AG (DE)  
FEATURES  
Source Location/Qualifiers  
1..7167  
/organism="synthetic construct"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)"  
BASE COUNT 2279 a 47 c 1050 g 3791 t  
ORIGIN

Query Match 8.5%; Score 52.8; DB 6; Length 7167;  
Best Local Similarity 46.3%; Pred. No. 2.7;  
Matches 174; Conservative 0; Mismatches 202; Indels 0; Gaps 0;  
Qy 234 TTTATTTATTTATTTGCTCAAGAGATGCTATTTTATGATGCAATATTTACTTTTAT 293  
Db 709 TATATTTGTTAGTTTATGAGATTTTAAAGATAGATTTTAAATATTTTAAATTTTAA 768  
Qy 294 TGCCTTACGCGCATGATTTTATACGTGATCAAGTGTGTCATTTAGCATACAAAGT 353  
Db 769 AGATTTTATGTTTATGATTTTATATATTTTAAATGTTTATGAGGTTTATGAAT 828  
Qy 354 TTTTAAATTTAAATGAGTAAATTTAATATATAAAGGTTGAAGAAATGTTGCTTC 413  
Db 829 ATTAATGTTTATGAGTAAATTTATATTTTAAATTTTAAAGTTTGAATGTTTAAATTT 888  
Qy 414 TCAACAAACACTTGTGACTTTTACGTAGTTTATGTTCAAGTAAATGTTATAGTATTT 473

Db 889 TGTGTTATAATGTTTATATGTTTATAGGTTTTCAGATGATTTGTGTTAAGTTTATTA 948  
Qy 474 TCCATATATGATTTAATAAAGAGATTTTCAACCTTTTATACCATATATATGCAATGCC 533  
Db 949 ATATTTGCGGTTTATTTAGCATTTTATAGTTTATTAATTTTATAGGTTTATTTAGATTA 1008  
Qy 534 TTCAGGTGTGCGAATTTATATGCTTAAATACAGAAATGACTTGTTTATATACGCTTAT 593  
Db 1009 TTTTATGTTTATGAGAAATTTATATTTTATGATATATTTATGTTTTCGATTAATTTATG 1068  
Qy 594 TCCATATGCAATTTAT 609  
Db 1069 TTTTATGAGATTTT 1084

RESULT 22  
LOCUS AE010546 11258 bp DNA linear BCT 25-MAR-2002  
DEFINITION Fusobacterium nucleatum subsp. nucleatum ATCC 25586, section 88 of 197 of the complete genome.  
ACCESSION AE010546 AE009951  
VERSION AE010546.1 GI:19713801  
KEYWORDS  
SOURCE Fusobacterium nucleatum subsp. nucleatum ATCC 25586  
ORGANISM Fusobacterium nucleatum subsp. nucleatum ATCC 25586  
REFERENCE 1 (bases 1 to 11258)  
AUTHORS Kaparatil,V., Anderson,I., Ivanova,N., Reznik,G., Los,T.,  
Lykidis,A., Bhattacharyya,A., Bartman,A., Gardner,W., Grechkin,G.,  
Zhu,L., Vasileva,O., Chu,L., Kogan,Y., Chaga,O., Goldsman,E.,  
Bernal,A., Larsen,N., D'Souza,M., Malunas,T., Pusch,G.D., Haselkorn,R.,  
Haselkorn,R., Fonstein,M., Kyripides,N. and Overbeek,R.  
TITLE Genome sequence and analysis of the oral bacterium Fusobacterium  
nucleatum strain ATCC 25586  
JOURNAL J. Bacteriol. 184 (7), 2005-2018 (2002)  
MEDLINE 21886394  
PUBMED 11889109  
REFERENCE 2 (bases 1 to 11258)  
AUTHORS Kaparatil,V., Anderson,I., Ivanova,N., Reznik,G., Los,T.,  
Lykidis,A., Bhattacharyya,A., Bartman,A., Gardner,W., Grechkin,G.,  
Zhu,L., Chu,L., Kogan,Y., Chaga,O., Goldsman,E., Bernal,A.,  
Larsen,N., D'Souza,M., Malunas,T., Pusch,G.D., Haselkorn,R.,  
Fonstein,M., Kyripides,N. and Overbeek,R.  
TITLE Direct Submission  
JOURNAL Submitted (13-FEB-2002) Integrated Genomics, 2201 W. Campbell Park  
Drive, Chicago, IL 60612, USA  
FEATURES  
Source Location/Qualifiers  
1..11258  
/organism="Fusobacterium nucleatum subsp. nucleatum ATCC  
25586"  
/mol\_type="genomic DNA"  
/strain="ATCC 25586"  
/sub\_species="nucleatum"  
/db\_xref="ATCC:25586"  
/db\_xref="taxon:190304"  
/complement(582..920)  
/gene="FN0337"  
complement(582..920)  
/gene="FN0337"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AA04541.1"  
/db\_xref="GI:19713802"  
/translation="MKKGNMKRFQNAATMEYNASKIDLVRDVNTNGIFPAIEFFENT  
KQIKRVETLVPVIEIQKNILKFSVQNNNNQSILTLILEINQVADKRSVINIS  
NEIDNNITLN"  
1169..1513  
/gene="FN0338"  
1169..1513  
/gene="FN0338"  
/codon\_start=1



gene  
CDS  
/transl\_table=11  
/product="Hypothetical protein"  
/protein\_id="AAL94542.1"  
/db\_xref="GI:19713803"  
/translation="MGSNFEKINQKVELGLVLEISEKDVAKAFIAYIFPPIFMI  
LGYIADOLGSEMSQIFSGFLGVGFIPLALVDRPFAKTIIDEIKVAIVEKIDPN  
ACTINAESECPF"  
1531..2004  
/gene="FN0339"  
1531..2004  
/gene="FN0339"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein"  
/protein\_id="AAL94543.1"  
/db\_xref="GI:19713804"  
/translation="MGMYGSYSLEKELEKLNKNDPDDIESVETLIDIKSWQAIQ  
YLGGDICEIGPLGVVPMVAENAIIDCSEGVFIYITBOIKKAVDALPPLTKEDLL  
EKVNSSEMLDEBVPYIVEDDEKEFEDVYSVLEIKKPYKKNIKEKALAIIPYS"  
complement(2084..3394)  
/gene="FN0340"  
complement(2084..3394)  
/gene="FN0340"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical membrane-spanning Protein"  
/protein\_id="AAL94544.1"  
/db\_xref="GI:19713805"  
/translation="MENKKYPTSVIIKELACSLIGIFLFPVPTLNGKSTIPLDHIV  
FLKIPYKEVYGLVLIIGVLPFYKKTWNKTSIVSFLKILALPFLMFLNNG  
PEFLNKDVIPIFNKIVIPVTIVVGSIFLSLISYLMGFVFWPPIKPIKMT  
PGRSADAVASVGSYSLALITNRYKGGKVTNKEAVIATGSEVSATPMVIAKTI  
LDLMNMLYFRLIYIVPLVTAIRARIYPIPNKSDAFBNQKDEVDIKREKXVA  
FNEGEVCTNSGSIENVIINLKSGIMAFNGBSLAIIGLIVLANHTPIEDVIGY  
LVYPTLLSGEESPLTAKALAGITEMFLPAVITKLSFEVKMLVAITCVSEVILFS  
ASIPCMATDIPISFKDYLIIFERFVVASIVSLIYLIVKVM"  
complement(3499..4827)  
/gene="FN0341"  
complement(3499..4827)  
/gene="FN0341"  
/codon\_start=1  
/transl\_table=11  
/product="transport protein"  
/protein\_id="AAL94545.1"  
/db\_xref="GI:19713806"  
/translation="MILNPVVISVIMTVICLLKLNVLAFISALVAGVTAGPIN  
DIMVGLITGMNGNETALSYILGLTAVAINSTGVSIVSKIASVAVGKKVLLVLI  
AFACFSOHLIFVHIAFTIPIPLKLMNDLDRAMACSLFGLKAPYITLPGV  
GAIFHSITAGETNNGMTVAQGDVWKSITWIGLPMIIGLLAFVSNKDRYKDLPL  
IGIERVADKMEAKHMLALIASAFITQALATVIRIKDGSILGAVVALLVNLIF  
GVYKKNIDEPFINGVGLMGLIAFTMLVAGAGNIIRGTGAINELVDSIHGILGSKA  
IGVSWMLIGLITWIGISFSTIVVAIYIPLCIKGVSSSIVLAAALALGDA  
GSPADSTLGPISGLNADQDHIMDTCPFTYHNILLIGFIMGFL"  
complement(4868..5371)  
/gene="FN0342"  
complement(4868..5371)  
/gene="FN0342"  
/EC\_number="5.2.1.8"  
/codon\_start=1  
/transl\_table=11  
/product="Peptidyl-prolyl cis-trans isomerase"  
/protein\_id="AAL94546.1"  
/db\_xref="GI:19713807"  
/translation="MSLOAITKNGEINLNLFSVAPYTVLNFITLAKGYVNGIKF  
HRVILFVQGDPTGTAGGPGYQFGDEFKGVFNKKGLAMNAGPNTNGSQFTI  
THVPTMLNFKHTIFGEVVSQKQDVVDISIKQDITMNEVTIIGDTDRILIDNKEFYTO  
LKNFLKI"  
complement(5425..6099)  
/gene="FN0343"  
complement(5425..6099)  
/gene="FN0343"  
/codon\_start=1  
/transl\_table=11

gene  
CDS  
/product="Hypothetical protein"  
/protein\_id="AAL94547.1"  
/db\_xref="GI:19713808"  
/translation="MEIERKMDIKIKIEDFIKYAKEVCLQNLPLADNIKVDIKNODLY  
EVEIERKEVISANIVLSLDEEPLNLNKENKCAFQLEBPTIEKMKDNLKQYIK  
TOIKRREKNGSABEYVERFPFKYIKELKIKDLOKLNKILDKKEKMLDLSNAI  
OEVEOLEIDKQPIRAEFNRLSIQDKYQKEIETENKLLKKMYEITYGTNKEILL  
KAVNSQ"  
complement(6068..7207)  
/gene="FN0344"  
complement(6068..7207)  
/gene="FN0344"  
/EC\_number="2.1.1.-"  
/codon\_start=1  
/transl\_table=11  
/product="Methyltransferase"  
/protein\_id="AAL94548.1"  
/db\_xref="GI:19713809"  
/translation="MIFLASTMGLSIVKDECLALGPNIKVFDGRVFEFGDEKDLI  
KANVLRCSDRVFIKMAEFKALTYEELFONIKSINMDFIDEDGEFPISWVSYSKL  
YSKSDIORISKALVEKLEKREITLENALYSIKIOSKODFIYWLDSGSSITK  
RGYARQKRVAPIKETTLAALVYSKMWADVLDPGCGTGTIAEAMAIANIANAGAN  
RNFAEKMSIIEKNLMTDIRDAFNSBDSLKELTIASDIDERSIEIAKENSEAGVE  
DIIIEYVDFKNIESPAKGYVMIIVPYGEIMGEEDIEELYRDPGNFCCKKLAKMSY  
YIITSYDEFEKAFPKATKQRKLYNGSIKCYQYQFGDRKNQYKN"  
complement(7207..8202)  
/gene="FN0345"  
complement(7207..8202)  
/gene="FN0345"  
/codon\_start=1  
/transl\_table=11  
/product="permease"  
/protein\_id="AAL94549.1"  
/db\_xref="GI:19713810"  
/translation="MDMKKQDQEMTINSIFAILLEPIVLYLKKSINDILASLSI  
VFIVIFVILSLIYPIHISIKINDIYPTISEKVIITGNIANYLAKNIYTIMD  
EIDVFNFNVSNNNTNITKLSSILGSLIDWTIGTFLAFVLAFLILDKHLIKT  
LENNIKIIFGVKNTPYMKLSKDLIFGVSGKIIVSTIVGCVYVLLITGPPVA  
ALSAIILGIGMIPVGSITIGGIIAFLIILVAPIKITIIILVAITISQNLVNGITGPK  
ITGNKVGINTFWIIVSMIFGNLFGIYGMFLGAPIMSLIKFYDDLKKAQGGF"  
complement(8321..8713)  
/gene="FN0346"  
complement(8321..8713)  
/gene="FN0346"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein"  
/protein\_id="AAL94550.1"  
/db\_xref="GI:19713811"  
/translation="MKKTKYKIDLVLYVFLIFFSILIKIGSFKDVKGAKEIIV  
DGEIKVYPIQKEEKNIFVETNIGGCVQFDMNVRVYTSPLKIANVKGQFISPGF  
VITIPDLVAVKVGDESDSGIDPFVAR"  
complement(8700..9602)  
/gene="FN0347"  
complement(8700..9602)  
/gene="FN0347"  
/EC\_number="4.1.1.65"  
/codon\_start=1  
/transl\_table=11  
/product="phosphatidylserine decarboxylase"  
/protein\_id="AAL94551.1"  
/db\_xref="GI:19713812"  
/translation="MNFPEKIOYIRKNGEIKTEKVMGSGALFKPLIYNPFGKALHAV  
KRKLSDMYGRKMSKPSKREKIKSFVEMETIDMSERYRPIIDVASFNDPFBELKDA  
RKIILNENVAIVSPADGKILAIQNKEDVKEFVKSQKTLFEPNDKEIAKKYEDGTFV  
IVRLAPADVHPHFPPVDGEISEIKKILGYYSVTHAKITFRILCEKREKRALILTE  
KFQDIAMEDIGATVGGIVQYKTNSSVYKGEKRGYFLFGSGTCLIVEKKNKVIDKD  
IIENTONKIEFRIYGEKFGNEKN"  
complement(9664..11169)  
/gene="FN0348"  
complement(9664..11169)  
8.5%; Score 52.8; DB 1; Length 11258;

Query Match

Best Local Similarity	46.7%;	Pred. No. 2.4;
Matches 168;	Conservative 0;	Mismatches 192;
		Indels 0;
		Gaps 0;

QY	161	ATTGTGAAACAACTTTTCTACATATATTTTAAAGATCATTTGCGTTGAGTTTACAGAC	220
Db	7839	ATTTTGTATATTTGTGTAGTATTTACTTACAAAATTAAGTAAATTAATCTATTTTAT	7898
QY	221	AGTCAGTAAATGCTTTATTTTATTTATGTGTGCAGAAGATAGCTATTTTATGTAGTCAA	280
Db	7899	CCATATTTATGATATTAATATTTTTTTTGGCTAAATAGTTTGTCTATATTTTTCCCTGTTA	7958
QY	281	TATACTTTTATTTGCTGTTACGCGATGATTTCTACGTACAAAGTTGGTGATAT	340
Db	7959	TTATCAGCTTTTCTGGAATGTAGAGATTAATATCATTTAAAGTTTATATGATGAATTA	8018
QY	341	GGACTCAAAAGCTTTTAAATTTAAATGAGTAATTTAATATATTAATAAGCGTTGAAAG	400
Db	8019	TTTCTGGAATTTACTATCAATGATGATAAAAATTAATTAATATTAATAAACTATTGATA	8078
QY	401	GAATGTTGCTTCCACAAACACTGTTCGACTTTACGATGTTAGTTCAGTAATTG	460
Db	8079	AACCTAATGCGAAATATCATTTATTTTACTTTTCTTTAAATATTTAACTATTTGGCT	8138
QY	461	TTATAGGTATGTTCCATATATGATTTAATAAAGAGAGCTTTCAACTTTTATCATAT	520
Db	8139	CTTAATAAGATAGCTATATTAATTTGACATTAATTTGATCAAAAATTAATCTTTTCCATTTAT	8198

RESULT 23	
AP002091	
LOCUS	AP002091 106763 bp DNA linear PRI 30-OCT-2001
DEFINITION	Homo sapiens genomic DNA, chromosome 6q25.2, clone: C1TD-2058ii12,
	complete sequence.
ACCESSION	AP002091
VERSION	AP002091.2 GI:15208256
KEYWORDS	.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1 Shimizu,N. and Asakawa,S.  
Homo sapiens DNA chromosome 6 SEQUENCE  
Published Only in Database (2000)  
2 (bases 1 to 106763)  
Shimizu,N. and Asakawa,S.  
Direct Submission  
Submitted (22-MAY-2000) Nobuyoshi Shimizu, Keio University, School

```
COMMENT      On Aug 16, 2001 this sequence version replaced gi:8096474.
FEATURES
source      location/qualifiers
            1..106763
```

```
/organism="Homo sapiens"  
/mol_type="genomic DNA"  
/db_xref="taxon:9606"
```

```

/clone="CITD-2058112"
/clone_lib="CIT Approved Human Genomic Sperm Library D
148.499
repeat_region

```

```
repeat_region      843_-1125  
/evidence=not experimental  
/rpt_family="A1uSg"  
repeat_region      complement(1227_-2493)  
/evidence=not experimental  
/rpt_family="A1uSx"  
repeat_region      2713_-3024  
/evidence=not experimental  
/rpt_family="A1uSg"
```

repeat_region	3027..3333	/evidence=not experimental
	/rpt_family="MULTIG1"	
repeat_region	3465..3499	/evidence=not experimental
	/rpt_family="Ar rich"	
repeat_region	3611..3706	/evidence=not experimental
	/rpt_family="M5B5"	
repeat_region	4131..4260	/evidence=not experimental
	/rpt_family="L4260..4558"	
repeat_region	4558..4847	/evidence=not experimental
	/rpt_family="AluY"	
repeat_region	5046..5587	/evidence=not experimental
	/rpt_family="MIR"	
repeat_region	5587..5886	/evidence=not experimental
	/rpt_family="AluSc"	
repeat_region	6557..6640	/evidence=not experimental
	/rpt_family="LIME"	
repeat_region	6640..6750	/evidence=not experimental
	/rpt_family="LIME"	
repeat_region	6751..7043	/evidence=not experimental
	/rpt_family="AluSg"	
repeat_region	7043..7075	/evidence=not experimental
	/rpt_family="LIME"	
repeat_region	7075..7108	/evidence=not experimental
	/rpt_family="TTTTT n"	
repeat_region	7109..7388	/evidence=not experimental
	/rpt_family="AluY"	
repeat_region	7389..7827	/evidence=not experimental
	/rpt_family="LIME"	
repeat_region	8049..8251	/evidence=not experimental
	/rpt_family="M5TA"	
repeat_region	8251..8517	/evidence=not experimental
	/rpt_family="M5TA"	
repeat_region	8517..8784	/evidence=not experimental
	/rpt_family="Ar rich"	
repeat_region	9041..9209	/evidence=not experimental
	/rpt_family="MIR"	
repeat_region	9211..9244	/evidence=not experimental
	/rpt_family="TTC n"	
repeat_region	9244..9271	/evidence=not experimental
	/rpt_family="TA n"	
repeat_region	10630..11137	/evidence=not experimental
	/rpt_family="MIR"	
repeat_region	11137..11175	/evidence=not experimental
	/rpt_family="Ar rich"	
repeat_region	11607..11808	/evidence=not experimental
	/rpt_family="AluSg"	
repeat_region	11808..13357	/evidence=not experimental
	/rpt_family="AluY"	
repeat_region	13357..14159	/evidence=not experimental
	/rpt_family="MIR5B"	
repeat_region	14159..1530	/evidence=not experimental

[illegible]

JOURNAL Nature 419 (6906), 527-531 (2002)  
MEDLINE 22255708  
PUBMED 12368667  
REFERENCE 2 (bases 1 to 254050)  
AUTHORS Sanders, M., Hauser, H., Baker, S., Unwin, L., Mungall, K., Bertman, M., Pain, A., Hall, N., Bowman, S., Church, C., Quail, M. and Barrett, B.  
TITLE Direct Submission  
Submitted (20-SEP-2002) P.falciparum Genome Sequencing Consortium, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Camps, Hinxton, Cambridge CB10 1SA, UK  
COMMENT For more information about this sequence or the Malaria Project, see [http://www.sanger.ac.uk/Projects/P\\_falciparum](http://www.sanger.ac.uk/Projects/P_falciparum).  
FEATURES  
Source Location/Qualifiers  
1. 254050  
/organism="Plasmodium falciparum 3D7"  
/mol\_type="genomic DNA"  
/isolate="3D7"  
/db\_xref="taxon:36329"  
/chromosome="9"  
763. 2532  
/gene="PF11215w"  
763. 2532  
/gene="PF11215w"  
/note="Moderately similar to Mus musculus splicing factor 3a, subunit 3, 60kd TR:AAH09141 (EMBL:BC009141) (501 aa) fasta scores: E(): 1.7e-24, 30.756% id in 595 aa. No significant pfam hit: SMART hit to SMO0355, zinc finger/ProfileScan hit to P850171, Zn-finger (C2H2) in matrin and other RNA-binding proteins"  
/codon\_start=1  
/product="spliceosome-associated protein, putative"  
/protein\_id="CAD51929.1"  
/db\_xref="GI:23505148"  
/translation="MSSQLLEIQRVIRHEIRVIRKAIKALIDKYNKKKILVDYALNYVEKIOKSQLLEYNDKDKKEMKTLSEGTGEDKDKIDKNTYKTYIRDRHKTINIKTILKSSPLKLEALCKSKYQSFSPLEKGGYVDLNFYDFIMKKIKERISIMFKKEMKASNNRKRKKTGNNNNNNNNNTTANINFEKMDYVTFYHNTRFIYIPRYCKRYNEYSYVENLNTMIISPEKINIVAVDCTYKEEDVYFNFKNEIKGMEKYTYDLDEYCNINNKLVASGEGYOSLYCKRYDNDIKYLNKKYTVGLEIEIKNIEKEDMLARCEVLIQLYANTINKTIQRTONIKQKAYTIDETQKKYKEBEDERKRRKETHILSLMDNNNELNNQNLLOLENNNISISQSDSDSMYSINNEQOKDLDEBEDDNDNKPITYNPLNPLGFNDKPYPYWKGLSKYCEICGNYSYFGNATFEKGFYEWRSFGMKCLIKIPNTLIPKETIKIDALNLYEKLKKEKTONIFKPOEYVCEDEKGVNNIKAYDILKROGL"  
join(4370). 4451.4555. 4604.4733. 4964.5141. 5186. 5417. 5444.5735. 5773.5790. 5825)  
/gene="PF1120w"  
/note="1 probable transmembrane helix predicted for PF11220w by TMHMM2.0 at aa 74-96"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAD51930.1"  
/db\_xref="GI:23505149"  
/translation="METFNHIDIKKYREDILYLRNTNVLFKMKQVDSIATLSK EOKICGLTISNNINENYTYITVTKKHITTYIVYVPLCLICISFTIYKLIIS CQVLTNHRMLRIKYSNNKLYMCVRLKTKKKKSTQHRPCLFEPNHFVIFSNFKNY MREPLINMC"  
join(6108). 6182.6392. 6547.6648. 6699.6851. 6903. 7009. 7062.7151. 7204.7343. 7424.7603. 7710.7795. 7853)  
/gene="PF11225w"  
join(6108). 6182.6392. 6547.6648. 6699.6851. 6903. 7009. 7062.7151. 7204.7343. 7424.7603. 7710.7795. 7853)  
/gene="PF11225w"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAD51931.1"  
/db\_xref="GI:23505150"  
/translation="MNEKKNIGISVHKEDNLSLTHQKEDLLIDYICDLKEEKERS TEKNTKGGDFVKSVDNRKADYAVRNNRKMHLHRLDENDEPITDILIEKKKIGTVR ITNPLPIERDELDKNYMKWLKTERLMRPNERLLEYEQMEKYLESIKILTHDIKQV

gene complement(8468. .9508)  
/gene="PF11230c"  
complement(8468. .9508)  
/gene="PF11230c"  
/note="asparagine-rich  
Signal anchor predicted for PF11230c by SignalP 2.0 HMM (signal peptide probability 0.001, signal anchor probability 0.995) with cleavage site probability 0.001 between residues 27 and 28  
4 probable transmembrane helices predicted for PF11230c by TMHMM2.0 at aa 12-34, 38-56, 63-82 and 97-116"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAD51932.1"  
/db\_xref="GI:23505151"  
/translation="MEERKWSYIMVSYIFLYCYPPELYSYEILFEPQIKDIYVPC LFTFSLFSIYFCGKGLKYLCTPLITISLIEIYEYNNIIPKDYDEFLFLVRL FSAVNLVPLFLHKNLKNIKDEKLNDDKOSTIOSVAPPTTENNHNNNNNNN IHNHNHNHNHNHNHNHNNDPCNNNDPCNNNYKQNDPLKSTYYSNNLPKEN IPRVHQTGIIITSTLPQENIYUAKHSIPFEKIIINKDTNISHMTEMQNDSPNH QVNNLSSNNINDKLTPEITTHNLNNNAVTNHIIONNHVPNTYEHPRGLQYYPG NFHTNNNGNNIO"  
join(11425. 12434.12559. 12655.12774. 12908)  
/gene="PF11235w"  
join(11425. 12434.12559. 12655.12774. 12908)  
/gene="PF11235w"  
/note="No pfam hit; ScanProsite hit to P500430; Interpro hit to Ton-B dependant receptor protein"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAD51933.1"  
/db\_xref="GI:23505152"  
/translation="WTDINNNKSKLVKKNKKNKNNKNNKNNKNTLLVSENDYKKE NKKKKHKNLQVYKIHNNNNNNNHKHDINDEOSNNHLYVNNRTKKKKNELIDKWI NKKKKKKKKKKKNTNTTIDYNDNNINDEEKKKKKYPPIISVIREVNNKIH IHDNIEYKTYDADLENNKOKYKKKKKINSPEIIVNSLERNVYTNVNSAVON KLNQTKVNNIYHQGYKOKKKKNNVNSIITIKLYKPKNNKTIADGCGEATART LDGYTISFPLIQNHVYTCNTQLPLNNSYDCLFSLSLANTWDPKITFESVRL KKGATVLIASVSRFTNYKAPIKFNNVNGRLSKINLDFEYVFEKQEDVISS INERIKRVSLPLPCYKAR"  
complement(13154. 14893)  
/gene="PF11240c"  
complement(13154. 14893)  
/gene="PF11240c"  
/EC\_number="6.1.1.15"  
/note="Similarities to prolyl-t-RNA synthases of bacterial origin; N-terminal signal peptide. Similar to Lactococcus lactis prolyl-tRNA synthetase pros or 112127 SWALL:Q9CDT4 (EMBL:AE006441) (616 aa) fasta scores: E(): 1.6e-09, 35.21% id in 213 aa, and to Escherichia coli prolyl-tRNA synthetase pros or dtpa or b0194 SWALL:GYP\_ECOLI (SWALL:P16659) (572 aa) fasta scores: E(): 7.3e-08, 31.91% id in 235 aa. Pfam match to PF00587, tRNA synthetase class II (G, H, P, S and T)  
Signal peptide predicted for PF11240c by SignalP 2.0 HMM (signal peptide probability 0.720, signal anchor probability 0.018) with cleavage site probability 0.591 between residues 18 and 19"  
/codon\_start=1  
/product="prolyl-t-RNA synthase, putative"  
/protein\_id="CAD51934.1"  
/db\_xref="GI:23505153"  
/translation="MLIKIILILGLFHHVILSLKLVNKKRSNVFVNKKILKNVP KKKYRILSNKQSYIGSNYIFNRIFFNNINSNGKISELORANYIVANGIYVILP GFRVINKIIDPLNKELEKNSHMSLSIILAKKLNISDSKLYSDEFLVYKOKOK NDKSSESLKKKIEDGYILSPICESSLSLINOIVENTITCLPLIHQYVKKFPK EKPEKSTFKSKETIMDQGSFHSNEKCLNTEYKCEYKANIPEELKLSFNITKKK KDKNNALSHSRFOVLSRDEGKREKAHIFKIGDYISNLDIKYDKKKNKNTLMGSG IGIRLLYFLIEFVYEEGKIKLPOQVAFSVYVLIQNOQKRSYATSKSLNMYKDSM EKGNNISQCNQNDQDDECGSSTFNSNHNDDKVNKDDTLNNDNTLNSNNTLNHD NGVNIINTLYSPILTSNDIEYIILTLMLFNFPKNNQVDIYVDDTDLHLSRLKHCDLIG

APNRIIINSLDKKVKVPTNFYNYINDSNILNKLYTFQNIITYEYKHPQSEKKI  
 MTIRELFYKPKRI"  
 complement(join(15848..16581,16944..17610))  
 /gene="PF11245c"  
 complement(join(15848..16581,16944..17610))  
 /gene="PF11245c"  
 /note="Plasmodium falciparum protein phosphatase-beta  
 TR:015920 (EMBL:U89025) (466 aa). fasta scores: E():  
 1.9e-180, 99.142% id in 466 aa. pfam match to pf000149,  
 Ser/Thr protein phosphatase; SMART hit to SM00156, Protein  
 phosphatase 2A homologues, catalytic domain."  
 /codon\_start=1  
 /product="Protein phosphatase-beta"  
 /protein\_id="CAD51935.1"  
 /db\_xref="GI:23505154"  
 /translation="MKCELSGRDIIENNVTNKNLKRKNKVGDLTYAEKPKDT  
 SLIEEDKDESELDCCGSEKKYKISNEENINTEPNPKENYIYNLSYKELSDS  
 EVSSENYNKKIKKNVNNPINNONSIEIYTSINDEKAGVINVTAGQIHGQFDILE  
 VDEMIKLLKELKIEYKLMCDLIDIKNEKNCVINVVTAGQIHGQFDILE  
 LEHIGLPPDVNVLFLGIVYDRGVYSCFCIACPKIKYPERVTLIGNHSRQITK  
 VYGFDECIKRYNNNNIYWKYITDYPDYLPLAIIINDEFCHGGISLQIOTIDINK  
 LDREKEIPQDGAICDLMSDPAGEDEITEGKSPRGAGVLPSEKTESFLRNKLS  
 CIGRAHQVQDFQWMMNDKVTIISAPNYCYRCNCASLMLVDFEMEKDPTFTNTP  
 LRANAQALRRNVQYML"  
 join(19483..19560,19785..19859,19939..20005,20102..20182,  
 20291..20343)  
 /gene="PF11250w"  
 join(19483..19560,19785..19859,19939..20005,20102..20182,  
 20291..20343)  
 CDS  
 gene  
 Query Match 8.5%; Score 52.6; DB 3; Length 254050;  
 Best Local Similarity 44.4%; Pred. No. 1.1;  
 Matches 211; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

QY 135 TATTATTAGCGCGGATGAGCGATTTTGAACAATTTCTACTATTATTACAGA 194  
 |||||  
 Db 237780 TGTATTATTACTTATTCATTATTATTATTATTATTATTATTATTATA 237839  
 |||||  
 QY 195 ATCATTGGGTTGAGGTACAGCAGCAGATGCTTATTATTATTATTATTGCTCA 254  
 |||||  
 Db 237840 TATATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 237899  
 |||||  
 QY 255 AGAGATAGCTATTATTATGATGCCATATTACTTTTATGCTGATGCGATGATTC 314  
 |||||  
 Db 237900 TTATCTTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 237959  
 |||||  
 QY 315 ATTACGTGACAAAGTTGTCATTTATGACTCAAAAGCTTTTAAATTAAATGAGCTA 374  
 |||||  
 Db 237960 ATTACTTATTTATCTTATTATTATTATTATTATTATTATTATTATTATT 238019  
 |||||  
 QY 375 ATTATATATATAAAGGCTGAAAGGATGTTGCTTCAACAACACTGTCGACT 434  
 |||||  
 Db 238020 ATTATATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 238079  
 |||||  
 QY 435 TTACGTAGTTAGTTCAAGTATGTTATAGTATGTTCCATATATGATATAAAG 494  
 |||||  
 Db 238080 ATTAAATATTATTATTATTATTATAGTATGATTTTATTATTATTATTATA 238139  
 |||||  
 QY 495 AAGAGTTTCAACTTTTACCATTAATTATGCAAGTCTTACGGTGGCAGATTAAT 554  
 |||||  
 Db 238140 TTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 238199  
 |||||  
 QY 555 GCTTATACAGGATAGTACTGTTTATATACGTAATTCATGACATTAAT 609  
 |||||  
 Db 238200 TTATATTATTAATTATTATTATTATTATTATTAATTTATTATTATT 238254  
 |||||

RESULT 25  
 AL390756 99003 bp DNA linear HMG 13-JUN-2001  
 LOCUS Homo sapiens chromosome 1 clone RP11-378A4, \*\*\* SEQUENCING IN  
 DEFINITION PROGRESS \*\*\* 29 unordered pieces.  
 ACCESSION AL390756  
 VERSION AL390756.2 GI:9931005

# KEYWORDS SOURCE ORGANISM

HTG; HTGS\_PHASE1; HTGS\_CANCELLED.  
 Homo sapiens (human)  
 Homo sapiens

## REFERENCE

McLay, K.  
 Direct Submission  
 Submitted (12-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 requests: clonerequests@sanger.ac.uk  
 On Aug 27, 2000 this sequence version replaced gi:9581725.

## COMMENT

----- Genome Center  
 Center: Sanger Centre  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquerry@sanger.ac.uk  
 Project Information  
 Center project name: ba378A4  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Sequencing vector: plasmid; L08752; 100% of reads  
 Chemistry: Dye-terminator; BT-ampersham; 26% of reads  
 Dye-terminator Big Dye; 73% of reads  
 Consensus quality: 84348 bases at least Q40  
 Consensus quality: 89481 bases at least Q30  
 Consensus quality: 92139 bases at least Q20  
 Insert size: 96203; sum-of-contigs  
 Insert size: 188174; 2.5% error; agarose-fp  
 Quality coverage: 2.40x in Q20 bases; sum-of-contigs Quality  
 coverage: 1.82x in Q20 bases; agarose-fp  
 -----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 29 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 2813: contig of 2813 bp in length  
 \* 2814 2913: gap of 100 bp  
 \* 2914 7380: contig of 4467 bp in length  
 \* 7381 7481: gap of 100 bp  
 \* 7481 11099: contig of 3619 bp in length  
 \* 11100 11199: gap of 100 bp  
 \* 11200 13897: contig of 2698 bp in length  
 \* 13898 13997: gap of 100 bp  
 \* 13998 16453: contig of 2456 bp in length  
 \* 16454 16554: gap of 100 bp  
 \* 16554 19055: contig of 2502 bp in length  
 \* 19056 19155: gap of 100 bp  
 \* 19155 21888: contig of 2733 bp in length  
 \* 21889 21988: gap of 100 bp  
 \* 21989 25542: contig of 3554 bp in length  
 \* 25543 25642: gap of 100 bp  
 \* 25642 30331: contig of 4699 bp in length  
 \* 30332 30431: gap of 100 bp  
 \* 30432 32487: contig of 2056 bp in length  
 \* 32488 32587: gap of 100 bp  
 \* 32588 36034: contig of 3447 bp in length  
 \* 36035 36134: gap of 100 bp  
 \* 36135 38214: contig of 2080 bp in length  
 \* 38215 38314: gap of 100 bp  
 \* 38315 40332: contig of 2018 bp in length  
 \* 40333 40432: gap of 100 bp  
 \* 40433 43002: contig of 2570 bp in length  
 \* 43003 43102: gap of 100 bp  
 \* 43103 48312: contig of 5110 bp in length  
 \* 48313 48312: gap of 100 bp  
 \* 48313 52476: contig of 4164 bp in length  
 \* 52477 52576: gap of 100 bp  
 \* 52577 54653: contig of 2077 bp in length















```

BASE COUNT      69938 a 38381 c 38675 g 70078 t      807 others
ORIGIN
/misc_feature   /note="assembly_fragment:02193
                 fragment_chain:1"
                 38739..52285
                 /note="assembly_fragment:02147
                 fragment_chain:1"
                 52386..66020
                 /note="assembly_fragment:00176
                 fragment_chain:2"
                 66121..138641
                 /note="assembly_fragment:01542
                 fragment_chain:2"
                 138742..145960
                 /note="assembly_fragment:01549
                 fragment_chain:3"
                 146061..190195
                 /note="assembly_fragment:02266
                 fragment_chain:3"
                 190296..193436
                 /note="assembly_fragment:00691.0"
                 193537..217879
                 /note="assembly_fragment:00862.0"
                 217880..38380
misc_feature     /note="assembly_fragment:00862.0"
                 38381..38674
misc_feature     /note="assembly_fragment:00862.0"
                 38675..70077
misc_feature     /note="assembly_fragment:00862.0"
                 70078..806
misc_feature     /note="assembly_fragment:00862.0"
                 807..807

```

Query Match	8.4%;	Score 52;	DB 2;	Length 217879;
Best Local Similarity	44.6%;	Pred. No. 1.5;		
Matches 205;	Conservative	0;	Mismatches 255;	Indels 0;
				Gaps 0;

Oy	162	TTTTGAACAAGATTTTCACGATAATTTTTTAAACAGAATCATTTGGSTTTGAGGTTCACGCACA	221
Dd	39578	TTTTTAATAATTATTTACCTTTTATTTTTTAAABRIGATTTAATATTTATATATATTC	395119
Oy	222	GTCAGTATATGCTTTATTTATTTATGTGTCCAGAAGATAGCTATTTTATTTGATGCCAAT	281
Dd	39518	TAAATATTTTATTTTATTTTATTTTATTTATGTTTATATATATATATATTTTATATATATTT	39459
Oy	282	ATTACTTTTATTTGCTGTACCGCAGCATTTCAATTACGTTTCAAGTGTGGCATTTANG	341
Dd	39458	ATTA	39399
Oy	342	GACTACAAAGGTTTTTAAATTTAAATGAGATAATTTAAATATATATATAAAGGTTGGAAGS	401
Dd	39398	TATTTGTAATTTTTTATTTACTTCACATCTTTATTTATTTTTTTAAAGATTAATTTAATA	39339
Oy	402	AATGTTGCTTTCACAAACACTTGTCAGCTTTTAGTGATTTAGTTCAAGTAATTTGT	461
Dd	39338	ATAATTTACTTTCTTAACCTAATTTGTATTATTTTATTTATATTTATATGTGATATATATATA	39279
Oy	462	TATATGATATGTTCCATATATATGATTTATATAAAGAGCTTTTCAAACCTTTTACCATTATA	521
Dd	39278	TA	39219
Oy	522	TTATGCAAGTCTTTCAGGTGTGCAGATTTATATGCTTTAATACAGAAATGTACTGTTTT	581
Dd	39218	TATATATATATACACATATATAAATATATAGAAAAATTTTATCTTTATTTACATTAATT	39159
Oy	582	ATATACGCTAATTTCTATGACATTTATTTGACATGTCAGAT	621
Dd	39158	ATTTACACTACACTCACTATTTTATATATATATACATTAAT	39119

RESULT	33
AX345066	
LOCUS	5523 bp DNA
DEFINITION	Sequence 137 from Patent WO0200928.
ACCESSION	AX345066
VERSION	AX345066.1 GI:18492952
KEYWORDS	.
SOURCE	synthetic construct
ORGANISM	synthetic construct
	artificial sequences.
REFERENCE	1
AUTHORS	Olek,A., Piepenbrock,C. and Berlin,K.

TITLE	JOURNAL	FEATURES	source
Diagnosis of diseases associated with the immune system	Patent: WO 0200928-A 137 03-JAN-2002; Epigenomics AG (DE)	Location/Qualifiers	1..5523
		/organism="synthetic construct"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:32630"	
		/note="chemically created genomic DNA (Homo sapiens)"	
BASE COUNT	1378 a	119 c	1366 g 2660 t
ORIGIN			

Query Match	8.3%	Score 51.8	DB 6	Length 5523
Best Local Similarity	45.6%	Pred. No. 4.5		
Matches 182; Conservative	0	Mismatches 217	Indels 0	Gaps 0

Oy	185	TTTTTACGAGATCACTTCGCGTTTAGGGTTACAGACAGACAGTACGTATATCTTATATATTT	244
Db	726	TTTTTATATAAAATTTTTTTTGTTTGAAGCTTTTGGATTTTAAAGTTATATAGTTTAAATAGTT	785
Oy	245	ATGTGTGTCAGAGATAGCTATTTTTATTTGATGCCAAATATTACTTTTATTTGCTGTAAACG	304
Db	786	ATGAGGTTTGAGTTTAAATAATTTTATGTGTTTGTTTTTATAGTTTAAATTTTATATAG	845
Oy	305	CATGAGATTCATTAACGTGTACAGTGGTGCAATTATGCACTACACAAAGTTTTTAAATTTA	364
Db	846	AATTAAGATAGTTAGAGTTGAATGAGGGGTTAAGAGATCGTTATTTTGTGTTTTTAAATAT	905
Oy	365	AATGAGATTAATTAATTAATTAATTAATAAAGGCTGGAAGGAATCTTGCTCTCAACAACAC	424
Db	906	ATAGATGAGGAGTTGTTATTTTAGAGCTGTAGAAATGTGTTATATGTTTTAGTTTTAGAT	965
Oy	425	TTGTTCGACTTTTACGTAGTTTACGTCAAGTAATTTGTAATAGTATGTTCCATATATGA	484
Db	966	TTTTTTTTTTTTTTTATAGTTTATTTAGTTATTTGTGATTTTAAAGAAAGATGTAATATA	1025
Oy	485	TTATTAAGAAGAGACTTTTCAACTTTTACCATTAATATATGCAAGTCTTCAGGTGG	544
Db	1026	GTAGTAAATATTTAAGAGTAATAAAGAGTGTATATGAAAGAAATTTATTTATATGAT	1085
Oy	545	CAGATTAATATCTTAATACAGAAATAGTACTGTTTTAT	583
Db	1086	TATATATATTAATTAATATAGCTTTATTTGTTGATAT	1124

```

RESULT 34
AY4242996/c
LOCUS          AY4242996          15575 bp    DNA          circular INV 15-MAR-2003
DEFINITION     Antheraea pernyi mitochondrial, complete genome.
ACCESSION      AY4242996
VERSION        AY4242996.1  GI:26974757
KEYWORDS
SOURCE
ORGANISM
Antheraea pernyi (Chinese oak silkmoth)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Bombycoidea; Saturniidae; Saturniinae; Saturniini; Antheraea.
1 (bases 1 to 15575)
Liu,Y., Xia,Y., Lu,C. and Xiang,Z.
Complete mitochondrial genome of Antheraea pernyi
Unpublished
2 (bases 1 to 15575)
Liu,Y., Xia,Y., Lu,C. and Xiang,Z.
Direct Submission
Submitted (23-FEB-2003) Key Sericultural Laboratory of the
Agricultural Ministry, Southwest Agricultural University, No.216
Tianshengqiao Road, Chongqing 400716, China
location/Qualifiers
1. 15575
/organism="Antheraea pernyi"
/organelle="mitochondrion"
/mol_type="genomic DNA"
/strain="Zao 1"
FEATURES
source

```

```

tRNA      /db_xref="taxon:7119"
          /dev_stage="pupa"
          /country="China: Henan"
          1..67
          /product="tRNA-Met"
tRNA      /product="tRNA-Ile"
          77..140
          complement(138..206)
tRNA      /product="tRNA-Gln"
          263..1276
          /gene="ND2"
          263..1276
          /gene="ND2"
          /codon_start=1
          /transl_table=5
          /product="NADH dehydrogenase subunit 2"
          /protein_id="AA053273.1"
          /db_xref="GI:28974758"
          /translation="MYPLNSKMKMFPIIPSTLIAISSNSMFCWIGLEINISLFI
          PLISSPNNLASASLKYFLQSIASINFLVILKMTFFKQFLDNFSPIMISSML
          KMGSPAPHPFNPVEGLSWENPILMTWQKITMILSYFNKFLYMTIIFMII
          GAIGLNOISLKLMAFSSINLGMMLPAIMISENLMTFETMYSFLSINCFEFL
          NMFPINOLFNNISPLIKINLINFSLGDLPPGLGFPKPMININFLNNKFKYLLFL
          FMMSLIIFFYIRITTYTCIMNTYKMKIKLSKNFLMNFSLISINGMISTE
          LFL"
tRNA      1285..1352
          /product="tRNA-Tyr"
          complement(1345..1405)
tRNA      /product="tRNA-Cys"
          complement(1425..1490)
tRNA      /product="tRNA-Tyr"
          <1497..3027
          /gene="COI"
          <1497..3027
          /gene="COI"
          /note="start codon undetermined: TAA stop codon is
          completed by the addition of 3' residues to the mRNA"
          /codon_start=1
          /transl_except=(pos:3027,aa:TERM)
          /transl_table=5
          /product="cytochrome c oxidase subunit I"
          /protein_id="AA053274.1"
          /db_xref="GI:28974759"
          /translation="RKMLYSTNHKDIGTYIFFGIWAGVGTSLIRAREIGTSGSL
          IGDDQIYNTIYTAHAFIMLFVWMPIMIGFGFNWLIPLMLGAPDAFPMKNWSTWLL
          PSLTLISSIYENGAGTGTVPPLSSNIHAGSSVDLAIFSLHAGISSIITAIN
          FTTIINNMNNLSFQDMPLEFWAGITAFALLSLPLVLAGATMLTDRNNTSFFD
          PAGGDPILYOHLEFMPFGHBEVYIILPQEGMISHISOESKKEFGCLSMYAMA
          IGLGRTVAHMTFTVGMIDIRAVFTSATMTIAYPTGKIKFSMLATTAGCQINSPS
          IIMSLGFVFLFTVGGITGVILANSSIDITLHPTIYVAHFHYVLSKGAFAIMGCFIH
          WYPLFTGLSLNPLFKIOPFIHFVGNLFFPQHFGLGMPRRYSDDPDSYLSWLLI
          SSLGYSILSLAVMLMLIIMESMIYORIALFSLNMFSSIEWYONLPPAHSHVNEPLIL
          SN"
tRNA      3028..3093
          /product="tRNA-Leu"
          3094..3775
          /gene="COII"
          3094..3775
          /gene="COII"
          /note="TAA stop codon is completed by the addition of 3'
          residues to the mRNA"
          /codon_start=1
          /transl_except=(pos:3775,aa:TERM)
          /transl_table=5
          /product="cytochrome c oxidase subunit II"
          /protein_id="AA053275.1"
          /db_xref="GI:28974760"
          /translation="MATWSNLNIQNGASPLMECIIFPHDHTLILIMITLYGYLMS
          LFFNSYINFLLEGOMTELIVTILPAITLIFALPRLLYLDELNNPLITKSGH
          OMWYSYSDPFNIEDSDYINPLNDLNNPRLVDNRIILPMNQIIRIMWTASDVI
          HSWTISLGIKVDANGRLNQTNFFINRGIIFYGQSEICGANHSFPIVIESISIKN
          FINMINVS"
          3776..3855

```

```

tRNA      /product="tRNA-Lys"
          3879..3915
          /product="tRNA-Asp"
          3952..4119
          /gene="ATP8"
          3952..4119
          /gene="ATP8"
          3952..4119
          /codon_start=1
          /transl_table=5
          /product="ATP synthase F0 subunit 8"
          /protein_id="AA053276.1"
          /db_xref="GI:28974761"
          /translation="MPQMPDINWFSFIFFIILFINIMNYIFENYINKNNKLIN
          NKMISFSMKW"
          4113..4790
          /gene="ATP6"
          4113..4790
          /gene="ATP6"
          /codon_start=1
          /transl_table=5
          /product="ATP synthase F0 subunit 6"
          /protein_id="AA053277.1"
          /db_xref="GI:28974762"
          /translation="MMGNLFSIPDPSNTIINTLSIMWLSTPLGLLFLPYSEFIPNRHF
          FWKFIILPKHNEFKTLKNNYFQSGTFIFSLSEFIFLFWLGLFPYIFSTSHLSL
          SLISPLMLSPMLXGWINNYQHFMIMIPQGTSLIPFWLLETISNIIIRPGTLAV
          RUTANMTAGHLMTLISNMSILPSYILMPLIFQIILLIESAVAVIQSVVIALST
          LYSSEVN"
          4790..5578
          /gene="COIII"
          4790..5578
          /gene="COIII"
          /codon_start=1
          /transl_table=5
          /product="cytochrome c oxidase subunit III"
          /protein_id="AA053278.1"
          /db_xref="GI:28974763"
          /translation="MKULYSHPYHLVDYSPWPLGAGVMTLVTLGVKPFNFSINLL
          IIGYITILITMYQWRDISRRCYQGGHTLIVTGLRMGMLFTVSELPFLSEFWAF
          FHSLSPNIDIGAMPPLSTIPNPFPHIPLINTIILISSGTVTAHAAHMLNNTOT
          NQSLFTIILGIVFTILOAVYELAPSIADSIYGAFFPMATFGHGLHVIIGTLPLIV
          CFIRHLSNHFSTHFGFEAAWYHFDVVMFLPLYSIYMWGN"
          5581..5646
          /product="tRNA-Gly"
          5647..6000
          /gene="ND3"
          5647..6000
          /gene="ND3"
          /codon_start=1
          /transl_table=5
          /product="NADH dehydrogenase subunit 3"
          /protein_id="AA053279.1"
          /db_xref="GI:28974764"
          /translation="MIIMMLIMITLILANIMFISITLSKSPFDRKSCPFEGCFD
          PKSSARIPSLHPFLITITLIFDVEIALIPPIYLFKTVNFPVWMTKTSFFIILLLL
          GYVHERNONMLTWIN"
          5993..6064
          /product="tRNA-Ala"
          6064..6128
          /product="tRNA-Arg"
          6129..6193
          /product="tRNA-Asn"
          6193..6260
          /product="tRNA-Ser"
          6262..6327
          /product="tRNA-Glu"
          complement(6338..6408)
          /product="tRNA-Phe"
          complement(6394..8149)
          /gene="ND5"
          complement(6394..8149)
          /note="TAA stop codon is completed by the addition of 3'

```

tRNA







```
* 4409 5165: contig of 757 bp in length
* 5166 5265: gap of 100 bp
* 5266 6031: contig of 766 bp in length
* 6032 6131: gap of 100 bp
* 6132 6896: contig of 765 bp in length
* 6897 7771: contig of 775 bp in length
* 7772 7872 8551: contig of 780 bp in length
* 8552 8752 9524: contig of 100 bp
* 9525 9625 10447: contig of 823 bp in length
* 10448 10547: gap of 100 bp
* 10548 11406: contig of 859 bp in length
* 11407 11507 12265: contig of 759 bp in length
* 11507 12266 12365: gap of 100 bp
* 12366 13122: contig of 757 bp in length
* 13123 13223 13980: contig of 758 bp in length
* 13981 14080: gap of 100 bp
* 14081 14922: contig of 842 bp in length
* 14923 15022: gap of 100 bp
* 15023 15805: contig of 783 bp in length
* 15806 15905: gap of 100 bp
* 15906 16664: contig of 759 bp in length
* 16665 16765 17537: contig of 773 bp in length
* 17538 17637: gap of 100 bp
* 17638 18409: contig of 772 bp in length
* 18410 18509: gap of 100 bp
* 18510 19278: contig of 769 bp in length
* 19279 19378: gap of 100 bp
* 19379 20137: contig of 753 bp in length
* 20138 20237: gap of 100 bp
* 20238 21012: contig of 775 bp in length
* 21013 21113 21892: contig of 780 bp in length
* 21113 21892: contig of 780 bp in length
* 21893 21992: gap of 100 bp
* 21993 22810: contig of 818 bp in length
* 22811 22910: gap of 100 bp
* 22911 23712: contig of 802 bp in length
* 23713 23812: gap of 100 bp
* 23813 24594: contig of 782 bp in length
* 24595 24694: gap of 100 bp
* 24695 25452: contig of 758 bp in length
* 25453 25552: gap of 100 bp
* 25553 26318: contig of 766 bp in length
* 26319 26418: gap of 100 bp
* 26419 27200: contig of 782 bp in length
* 27201 27300: gap of 100 bp
* 27301 28095: contig of 795 bp in length
* 28096 28195: gap of 100 bp
* 28196 28977: contig of 782 bp in length
* 28978 29077: gap of 100 bp
* 29078 29883: contig of 806 bp in length
* 29884 29983: gap of 100 bp
* 29984 30776: contig of 793 bp in length
* 30777 30876: gap of 100 bp
* 30877 31661: contig of 785 bp in length
* 31662 31761: gap of 100 bp
* 31762 32517: contig of 756 bp in length
* 32518 32617: gap of 100 bp
* 32618 33392: contig of 775 bp in length
* 33393 33492: gap of 100 bp
* 33493 34254: contig of 762 bp in length
* 34255 34354: gap of 100 bp
* 34355 35148: contig of 794 bp in length
* 35149 35248: gap of 100 bp
* 35249 36040: contig of 792 bp in length
* 36041 36140: gap of 100 bp
* 36141 36912: contig of 772 bp in length
```

```
* 36913 37012: gap of 100 bp
* 37013 37807: contig of 795 bp in length
* 37808 37907: gap of 100 bp
* 37908 38664: contig of 757 bp in length
* 38665 38764: gap of 100 bp
* 38765 39602: contig of 838 bp in length
* 39603 39703 40477: contig of 775 bp in length
* 39703 40577: gap of 100 bp
* 40578 41344: contig of 766 bp in length
* 41344 41443: gap of 100 bp
* 41444 42203: contig of 760 bp in length
* 42204 42304: gap of 100 bp
* 42304 43068: contig of 764 bp in length
* 43068 43168: gap of 100 bp
* 43168 43964: contig of 796 bp in length
* 43964 44063: gap of 100 bp
* 44064 44885: contig of 822 bp in length
* 44886 44986: gap of 100 bp
* 44986 45745: contig of 760 bp in length
* 45746 45846: gap of 100 bp
* 45846 46616: contig of 770 bp in length
* 46616 46716: gap of 100 bp
* 46716 47482: contig of 767 bp in length
* 47483 47583 48371: contig of 789 bp in length
* 48372 48472: gap of 100 bp
* 48472 49249: contig of 777 bp in length
* 49249 49349: gap of 100 bp
* 49349 50123: contig of 775 bp in length
* 50124 50223: gap of 100 bp
* 50224 51000: contig of 777 bp in length
* 51001 51100: gap of 100 bp
* 51101 51857: contig of 757 bp in length
* 51858 51958: gap of 100 bp
* 51958 52818: contig of 861 bp in length
* 52819 52919: gap of 100 bp
* 52919 53694: contig of 776 bp in length
* 53695 53795: gap of 100 bp
* 53795 54564: contig of 770 bp in length
* 54564 54664: gap of 100 bp
* 54665 55453: contig of 789 bp in length
* 55454 55554: gap of 100 bp
* 55554 56424: contig of 871 bp in length
* 56425 56524: gap of 100 bp
* 56525 57278: contig of 754 bp in length
* 57279 57379: gap of 100 bp
* 57379 58156: contig of 778 bp in length
* 58157 58257: gap of 100 bp
* 58257 59009: contig of 753 bp in length
* 59109 59110: gap of 100 bp
* 59110 59903: contig of 794 bp in length
* 59904 60003: gap of 100 bp
* 60004 60783: contig of 780 bp in length
* 60784 60883: gap of 100 bp

Query Match      8.3%; Score 51.6; DB 2; Length 63513;
Best Local Similarity 44.5%; Pred. No. 2.5;
Matches 186; Conservative 0; Mismatches 232; Indels 0; Gaps 0;

Qy 164 TTGAACAATTTTCTACTATATTTTTCACAGATCATTTGGTTGAGGTTACACACAGT 223
Db 38964 TTGTTTANGTTTNTTANATTTTTTTTATTTATTTAAATTTTTTTGGGGGAGTAA 39023

Qy 224 CAGTATATGCTTATTTATTTATTTATTTAGTCTCAGAGATAGCTATTTATTTGATGCAATAT 283
Db 39024 TATATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 39083

Qy 284 TACTTTTATTTGCTGATGACGATGATTCATTTACGATGATGATGATGATGATGATGATGAT 343
Db 39084 TTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 39143

Qy 344 CTACAAAGGTTTTTAAATTTAAATGAGATAATTTAATATATTAATAAGGTTGAAGGAA 403
```

Db 39144 TAAATTAATTTTGTATATATATTTTAAATTTTATATATTTTATTTTATATTA 39203

Qy 404 TGTTCCTTCGAACAACACTGTTCCAGCTTTACGATGTTAGTCAAGTAATGTTA 463

Db 39204 TATTTTATTTATTTTAAATTTTAAATATATATTAATTAATTAATTTATTTATTA 39263

Qy 464 TAGGATTTGTCATATATGATTAATAAGAGAGTTTCAACCTTTACCATATAT 523

Db 39264 TTAATATTTATTTATTTATTTTATTTTATTTTATTTTATTTATTTATTTATTTATTT 39323

Qy 524 ATGCAAGTCCTTCAGGTGCGAGATTAATGCTTAATACAGAAATGACTGTTT 581

Db 39324 AATTATATATTTTATTTATTTTATTTTATTTTATTTTATTTATTTATTTATTTATTT 39381

RESULT 39

EX004827/c 119555 bp DNA linear PRI 11-APR-2003

LOCUS Human DNA sequence from clone RP13-391G2 on chromosome X, complete

DEFINITION

ACCESSION BX004827

VERSION BX004827.18 GI:29824042

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

AUTHORS 1 (bases 1 to 119555)

TITLE Direct Submission

JOURNAL Submitted (11-APR-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: clones@wells.sanger.ac.uk

COMMENT On Apr 11, 2003 this sequence version replaced gi:29603318.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Web site: http://www.sanger.ac.uk

Contact: humquerry@sanger.ac.uk

-----

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMREP; Information on the WORMREP database can be found at http://www.sanger.ac.uk/Projects/C\_elegans/wormrep This sequence was generated from part of bacterial clone configs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/ChrX

RP13-391G2 is from the library RPCT-13.2 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm

VECTOR: pBACe3.6.

FEATURES

SOURCE

1..119555

Location/Qualifiers

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

BASE COUNT 30503 a 28003 c 27878 g 33171 t

ORIGIN

Query Match 8.3%; Score 51.6; DB 9; Length 119555; Best Local Similarity 48.8%; Pred. No. 2.1; Matches 168; Conservative 0; Mismatches 174; Indels 2; Gaps 1;

Qy 181 TATAATTTTACAGATCATTTTCGTTGAGGTTACAGACAGTCAGATATGCTTATTT 240

Db 86743 TATATATGTTATATATATTTTATATGATATATATATATATATATATATATATAT 86684

Qy 241 ATTTATGTCGTCAGAGATAGCTATTTATTTATGCAATATTTACTTTTATGCTGTT 300

Db 86683 ATATATGTTTATATATATGATATTTTATATATATATATATATATATATATATAT 86624

Qy 301 ACGCATGATTTTCATTAAGTGTACAGATTTGTCATTTAGCATACAAAGTTTAA 360

Db 86623 TATATATGTTTATATATATTTTATATGATATATATATATATATATATATATATAT 86564

Qy 361 TTTAATGAGATTAATTTATATATATATATATATATATATATATATATATATATAT 86504

Db 86563 TACAATATGTTTGTATATATATTTATATATATATATATATATATATATATATATAT 86504

Qy 419 AAACACTGTCGACTTTTACGTTAGTGTACAGATATTTATGTTATGTTTCAT 478

Db 86503 ATATATTTTATACATATTTTATATATATATATATATATATATATATATATATATAT 86444

Qy 479 ATATGATTAATAAGAGAGATTTCAACCTTTACCATTTAT 522

Db 86443 TTAATGTTTAT 86400

RESULT 40

CPU88070 10807 bp DNA linear BCT 07-MAY-1999

LOCUS Chlamydomophila caviæ cds1, cds2, copen and sccl genes, complete cds.

DEFINITION

ACCESSION U88070

VERSION U88070.1 GI:2444072

KEYWORDS

SOURCE Chlamydomophila caviæ

ORGANISM Chlamydomophila caviæ

REFERENCE 1 (bases 1 to 10807) Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.

AUTHORS Hsiao,R.C., Pannekoek,Y., Ingerowski,E. and Bavoil,P.M.

TITLE Type III secretion genes identify a putative virulence locus of Chlamydia

JOURNAL Mol. Microbiol. 25 (2), 351-359 (1997)

MEDLINE 97426043

PUBMED 9282747

REFERENCE 2 (bases 1 to 10807) Hsiao,R.C., Pannekoek,Y., Ingerowski,E. and Bavoil,P.M.

AUTHORS Direct Submission

TITLE Submitted (03-FEB-1997) Clinical Sciences, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

COMMENT On Sep 30, 1997 this sequence version replaced gi:2358255.

----- Location/Qualifiers

1..10807

/organism="Chlamydomophila caviæ"

/mol\_type="genomic DNA"

/strain="GPIC"

/specific\_host="guinea pig"

/db\_xref="taxon:83557"

/note="Submitted as Chlamydia psittaci guinea pig inclusion conjunctivitis"

315..1397

/gene="cds1"

315..1397

/gene="cds1"

315..1397

/gene="cds1"

/function="cytoplasmic membrane component required for type III (contact-dependent) secretion"

gene

CDS

1..10807

/organism="Chlamydomophila caviæ"

/mol\_type="genomic DNA"

/strain="GPIC"

/specific\_host="guinea pig"

/db\_xref="taxon:83557"

/note="Submitted as Chlamydia psittaci guinea pig inclusion conjunctivitis"

315..1397

/gene="cds1"

315..1397

/gene="cds1"

315..1397

/gene="cds1"

/function="cytoplasmic membrane component required for type III (contact-dependent) secretion"



Qy 259 ATAGCTATTTTATGATGCCAATATTACTTTTATGCTGTACGGCATGGATTTCATTA 318  
Db 564 CTAAACCACTTCCCTT---CCCTACTGTGTGCTGTGGGATTGTGGTATTATGTGCGA 620  
Qy 319 CGGTACAGTTGTGTGCAATTATGACTACAAAGTTTAAATTAAATGAGTAATTT 378  
Db 621 TTCTGTGTGTAGGCCCACTTTTCTACAGAGGTTTAAAGCTTAATAAAGTTC 680  
Qy 379 AATAAATATAAGGTTGAAGGAATGTT 407  
Db 681 AATCCTATTGACAACCTTGAACAAAATTT 709

Search completed: October 4, 2003, 04:39:26  
Job time : 2061 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 3, 2003, 18:14:56 ; Search time 217 Seconds  
(without alignments)  
7737.560 Million cell updates/sec

Title: US-10-010-160-1

Perfect score: 622

Sequence: 1 atgcgcgatccagctaa.....aattatgcagtcgcagatc 622

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_190un03:\*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*
- 25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	622	100.0	622	24	ABK97884
2	54.8	8.8	8056	25	ABZ10100
3	54.8	8.8	9052	24	ABZ12145
4	54.4	8.6	8056	25	ABZ10246
5	53.6	8.6	1501	25	ABZ10188
6	53	8.5	945	19	AAK14127
7	53	8.5	1024	24	ABK6125
8	53	8.5	1086	18	AAK67941

9	52.8	8.5	7167	24	ABL32401	Human immune syste
10	52	8.4	5297	22	AA546514	Tumour suppressor
11	52	8.4	7128	24	ABL33559	Human immune syste
12	52	8.4	40324	24	ABOC7150	Human angiogenesis
13	51.8	8.3	5523	24	ABL32164	Human immune syste
14	51.8	8.3	73334	24	ABL92318	Chemically treated
15	51.8	8.3	73334	24	ABL34124	Human immune syste
16	51.2	8.2	12025	24	ABL33299	Human immune syste
17	51	8.2	6118	24	ABL33031	Human immune syste
18	51	8.2	7921	24	ABL33971	Human immune syste
19	50.6	8.1	6126	22	AA546574	Tumour suppressor
20	50.6	8.1	6126	24	ABL33831	Human immune syste
21	50.6	8.1	11047	22	AA545480	Chemically pretrea
22	50.6	8.1	11047	24	ABL33985	Human immune syste
23	50.6	8.1	11047	24	ABK28402	DNA transcription
24	50.2	8.1	9118	24	ABOC6989	Human angiogenesis
25	50.2	8.1	9543	24	ABN80048	Human chemically m
26	50.2	8.1	513445	22	AA161373	Soybean 318013 reg
27	49.8	8.0	5935	22	AA545426	Chemically pretrea
28	49.8	8.0	40862	24	ABL34072	Human immune syste
29	49.8	8.0	1038602	20	AAZ01425	Complete genome se
30	49.2	7.9	61020	22	AA546788	Tumour suppressor
31	48.8	7.8	5542	24	ABL34021	Human immune syste
32	48.8	7.8	5857	24	AA563347	Chemically pretrea
33	48.6	7.8	18855	24	ABL32610	Human immune syste
34	48.4	7.8	11787	24	ABL92243	Chemically treated
35	48.4	7.8	19345	24	ABN80017	Human chemically m
36	48.2	7.7	12138	24	ABK40034	Human immune syste
37	48.2	7.7	12138	24	ABL33629	Human immune syste
38	48.2	7.7	12590	24	ABL33920	Human immune syste
39	48.2	7.7	18512	24	ABL32977	Human immune syste
40	48.2	7.7	82952	24	ABN85766	Arabidopsis yellow
41	48	7.7	13427	24	ABL33937	Human immune syste
42	47.8	7.7	524	24	ABO34338	Oligonucleotide fo
43	47.8	7.7	524	24	ABO34339	Oligonucleotide fo
44	47.8	7.7	5126	24	ABL70493	Chemically treated
45	47.8	7.7	6310	24	ABL70321	Chemically treated

## ALIGNMENTS

RESULT 1	ABK97884	standard; DNA; 622 BP.
ID	ABK97884	
XX	ABK97884;	
AC		
XX		
DT	10-OCT-2002 (first entry)	
XX		
DE	DNA encoding Lawsonia intracellularis antigenic flhB protein.	
XX		
KW	Immunogen; flhB; flhR; ntrC; glhH; motB; clyC; ytfW; ytfH;	
KW	porcine; pig; avian; bird; porcine proliferative enteropathy; PPE;	
KW	intestinal adenomatosis complex; porcine intestinal adenomatosis;	
KW	PIA; necrotic enteritis; proliferative haemorrhagic enteropathy;	
KW	regional ileitis; haemorrhagic bowel syndrome; vaccine; antibacterial;	
KW	porcine proliferative enteritis; Campylobacter spp.-induced enteritis;	
KW	gene; ds.	
XX		
OS	Lawsonia intracellularis.	
XX		
PN	WO200238594-A1.	
XX		
PD	16-MAY-2002.	
XX		
PF	09-NOV-2001; 2001WO-AU01462.	
XX		
PR	10-NOV-2000; 2000AU-0001381.	
PR	17-NOV-2000; 2000US-249596P.	
XX		
PA	(AGRI-) AGRIC VICTORIA SERVICES PTY LTD.	
PA	(AUPO-) AUSTRALIAN PORK LTD.	

PA (PRIZ ) PFIZER PROD INC.

XX Rosey EL, King KW, Good RT, Strugnelli RA;

XX WPI, 2002-557448/59.

DR P-PSDB; ABG68910.

XX New immunogenic polypeptide comprising epitope of Lawsonia spp.

PT polypeptide such as flhb, flir, ntrc, glnh, moca, polypeptides, useful

PT in vaccines for treatment of porcine proliferative enteropathy in pigs

PS and birds

PS Claim 6; Page 99-100; 155pp; English.

CC The invention describes an isolated or recombinant immunogenic  
CC polypeptide (1) which comprises, mimics or cross-reacts with a B-cell or  
CC T-cell epitope of a Lawsonia spp. polypeptide such as flhb, flir, ntrc,  
CC glnh, moca, moca, tlyc, yfem or yfem polypeptides. (1) is useful for  
CC identifying whether or not a porcine or avian animal has suffered from a  
CC past infection, or is currently infected, with Lawsonia spp. or a  
CC microorganism that is immunologically cross-reactive with Lawsonia spp.  
CC Antibodies are useful for diagnosing infection of a porcine or avian  
CC animal by Lawsonia spp. or a microorganism that is immunologically  
CC cross-reactive with Lawsonia spp.. A nucleic acid encoding a Lawsonia  
CC spp. immunogen is useful as probes or primers for detecting Lawsonia  
CC or, related microorganism in a biological sample derived from a porcine or  
CC avian animal subject. (1) is preferably useful for vaccinating porcine  
CC animals against intestinal diseases collectively known as intestinal  
CC proliferative enteropathy (PPE), previously known as intestinal  
CC adenomatosis complex, porcine intestinal adenomatosis (PIA), necrotic  
CC enteritis, proliferative haemorrhagic enteropathy, regional ileitis,  
CC haemorrhagic bowel syndrome, porcine proliferative enteritis and  
CC Campylobacter spp.-induced enteritis. (1) is also useful in vaccines for  
CC the prophylaxis and treatment of PPE in birds. This sequence encodes a  
CC Lawsonia intracellularis immunogenic peptide used in the creation of a  
CC porcine vaccine described in the invention.

XX Sequence 622 BP; 189 A; 92 C; 123 G; 218 T; 0 other;

Query Match 100.0%; Score 622; DB 24; Length 622;  
Best Local Similarity 100.0%; Pred. No. 1.2e-135;  
Matches 622; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTGATGATCCAGTAAACAGAGAAAGCAACCCGAAACGTCAGGAAGCTGT 60  
DB 1 ATGCTGATGATCCAGTAAACAGAGAAAGCAACCCGAAACGTCAGGAAGCTGT 60  
QY 61 TCTGAAGGAGAGTGTCCCTTAATCAGAGAGTTACTTAAAGCTACTGACAGAGG 120  
DB 61 TCTGAAGGAGAGTGTCCCTTAATCAGAGAGTTACTTAAAGCTACTGACAGAGG 120  
QY 121 ATGCTGGGAGCTGTATTATTCAGGCGTAATGGAACGTCATTGGAACAATTTTCTAC 180  
DB 121 ATGCTGGGAGCTGTATTATTCAGGCGTAATGGAACGTCATTGGAACAATTTTCTAC 180  
QY 181 TATATTTTACAGATCAATTCGTTGAGTTAGAGCAAGTATGCTATATGCTTATTT 240  
DB 181 TATATTTTACAGATCAATTCGTTGAGTTAGAGCAAGTATGCTATATGCTTATTT 240  
QY 241 ATTTATGTTGCTCAAGAGATAGCTATTTTATGATGCCAATTTATCTTTTATTCGTT 300  
DB 241 ATTTATGTTGCTCAAGAGATAGCTATTTTATGATGCCAATTTATCTTTTATTCGTT 300  
QY 301 ACGGATGATTTTCAATTCAGTGTACAGAGTGTGATTAAGACATCAAAAGGTTTAA 360  
DB 301 ACGGATGATTTTCAATTCAGTGTACAGAGTGTGATTAAGACATCAAAAGGTTTAA 360  
QY 361 TTTAAATGAGATTTTAAATTAATTAATAAGGTTGAAAGAGATGTTGCTTCAACA 420  
DB 361 TTTAAATGAGATTTTAAATTAATTAATAAGGTTGAAAGAGATGTTGCTTCAACA 420  
QY 421 ACACTGTTGCACTTTTACGTAGTTAGTTCAAGTAATGTTATAGTATTTTCCATAT 480  
DB 421 ACACTGTTGCACTTTTACGTAGTTAGTTCAAGTAATGTTATAGTATTTTCCATAT 480

DB 421 ACACTGTTGCACTTTTACGTAGTTAGTTCAAGTAATGTTATAGTATTTTCCATAT 480  
QY 481 ATGATTTAAAGAGAGAGCTTTTCAACTTTTACCATTATATATGCAAGTCTTCAGGT 540  
DB 481 ATGATTTAAAGAGAGAGCTTTTCAACTTTTACCATTATATATGCAAGTCTTCAGGT 540  
QY 541 GTGGCAGATTTATGCTTTAATACAGAAATATCTGTTTATATAGCCTAATTCCTATG 600  
DB 541 GTGGCAGATTTATGCTTTAATACAGAAATATCTGTTTATATAGCCTAATTCCTATG 600  
QY 601 ACAATTTATGCGAGTCGAGATC 622  
DB 601 ACAATTTATGCGAGTCGAGATC 622

RESULT 2  
ABZ10100  
ID ABZ10100 standard; DNA; 8056 BP.

AC ABZ10100;

DT 16-JAN-2003 (first entry)

DE Haematopoietic cell proliferation disorder related DNA sequence #240.

KW Human; haematopoietic cell proliferation disorder; cytostatic;  
KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;  
KW cytosine methylation state; gene; ds.

OS Homo sapiens.

PN WO20027272-A2.

PD 03-OCT-2002.

PF 26-MAR-2002; 2002WO-EP03401.

PR 26-MAR-2001; 2001US-278333P.

PA (EPIG-) EPIGENOMICS AG.

PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;  
PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Liesche R, Lau B;  
PI Lehn A, Lipscher B, Meier S, Model F, Mueller V, Otto T;  
PI Pelet C, Schwabe I, Ziebarth H;

DR WPI, 2003-018942/01.

PT Detecting and differentiating between hematopoietic cell proliferative  
PT disorders, comprises contacting a target nucleic acid with a reagent  
PT that distinguishes between methylated and non-methylated CpG  
PT dinucleotides -

PS Claim 28; SEQ ID 240; 117pp; English.

XX The present invention describes a method for detecting and  
XX differentiating between haematopoietic cell proliferative disorders  
XX associated with at least 1 gene and/or their regulatory regions in a  
XX subject. The method comprises contacting a target nucleic acid in a  
XX biological sample obtained from the subject with at least 1 reagent,  
XX which distinguishes between methylated and non-methylated CpG  
XX dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118  
XX represent specifically claimed nucleotide sequences from the present  
XX invention. Oligonucleotides from the present invention can be used: for  
XX differentiating between healthy haematopoietic cells and proliferative  
XX disorder haematopoietic cells; for differentiating between acute  
XX lymphocytic leukaemia and acute myelogenous leukaemia; as probes for  
XX determining the cytosine methylation state and/or single nucleotide  
XX polymorphisms (SNPs) of haematopoietic cell proliferation disorder  
XX related sequences and their complements; and as primers for the  
XX amplification of haematopoietic cell proliferation disorder related  
XX DNA sequences. The nucleotide sequences from the present invention can  
XX also be used for detecting a predisposition to, differentiation between



CC subclasses, diagnosis, prognosis, treatment and/or monitoring of  
CC hematopoietic cell proliferative disorders. The present method enables  
CC a highly specific classification of hematopoietic cell proliferative  
CC disorders allowing for improved and informed treatment of patients.

XX Sequence 8056 BP; 3711 A; 371 C; 371 G; 3603 T; 0 other;

Query Match 8.8%; Score 54.8; DB 25; Length 8056;  
Best Local Similarity 47.2%; Pred. No. 0.0035;  
Matches 167; Conservative 0; Mismatches 187; Indels 0; Gaps 0;

QY 160 CATTGGAACAATTTCTACTATATTTTACAGAAATCTGGCTTGAGCTTACGCA 219  
DB 6351 CGTATAAATAAATTTTATTTATTTATTTATTTATTTATTTATTTATTTAT 6410  
QY 220 CAGTCAGTATATGCTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 279  
DB 6411 ATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 6470  
QY 280 ATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 339  
DB 6471 ATATATATATATCGTTTTTTTATTTATTTATTTATTTATTTATTTATTTATTTAT 6530  
QY 340 TGGACGTACAAAGCTTTTAAATTTAAATGAGTAAATTTATTTATTTATTTATTTATTTAT 399  
DB 6531 AACGTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 6590  
QY 400 GGAATGTTGCTGTCACAAACACTGTTGCACTTTTACGACTTTTACGACTTTTACGACTTT 459  
DB 6591 AAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 6650  
QY 460 GTATATAGTATTTGTCATATATGATTTATTTAAAGAGAGTTTTCAACTTTTAA 513  
DB 6651 AATTTTAAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 6704

RESULT 3  
AB212145  
ID ABL32145 standard; DNA; 9052 BP.

AC ABL32145;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Human immune system associated gene SEQ ID NO: 118.

KM Human; immune system disease; cytosine methylation; antiasthmatic;  
KM antiarteriosclerotic; anti-nausea; cytosine; nootropic;  
KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KM antineumatic; antiaortic; antidiabetic; antiporotic;  
KM antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KM gene; ds.

OS Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful

PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation  
XX  
XX Claim 1; SEQ ID NO 118; 32zp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.

XX Sequence 9052 BP; 2166 A; 112 C; 2104 G; 4670 T; 0 other;

Query Match 8.8%; Score 54.8; DB 24; Length 9052;  
Best Local Similarity 47.2%; Pred. No. 0.0036;  
Matches 167; Conservative 0; Mismatches 187; Indels 0; Gaps 0;

QY 138 TTAATGAGGCTATGGAGCTCATTTTGAACAATTTTCTACTATATTTTACAGATC 197  
DB 799 TTTTTCGATGAGAAAAGAGCTTTATTTATTTATTTATTTATTTATTTATTTATTTAT 858  
QY 198 ATTCGCTTGAAGTTACACACAGTCAGTATGCTTTATTTATTTATTTATTTATTTATTTAT 257  
DB 859 TTAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 918  
QY 258 GATAGCTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 317  
DB 919 TTTGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 978  
QY 318 ACGTGTACAGTTGCTGCTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 377  
DB 979 ATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1038  
QY 378 TAAATATATTTAAAGGTTGAAGAGATTTGCTTCAACAAACACTGTTGACCTTTT 437  
DB 1039 TGTGAATAATATAGATTTTAAATGATGATATTTGTAAGTATTTTATTTTATTTATTTAT 1098  
QY 438 ACGTATTTGTTCAAGTAAATGTTATAGTATTTGCTCATATATTTATTTAA 491  
DB 1099 AGGTAAAGTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1152

RESULT 4

AB210246  
ID AB210246 standard; DNA; 8056 BP.

XX AB210246;

XX 16-JAN-2003 (first entry)

XX Haematopoietic cell proliferation disorder related DNA sequence #386.

KM Human; haematopoietic cell proliferation disorder; cytosine;  
KM gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;  
KM cytosine methylation state; gene; ds.

OS Homo sapiens.

XX WO20027272-A2.

XX 03-OCT-2002.

XX 26-MAR-2002; 2002WO-EP03401.

XX 26-MAR-2001; 2001US-278333P.

XX (EPIC-) EPIGENOMICS AG.

XX Berlin K, Braun A, Distler J, Guelig D, Howe A, Mueller J;

XX Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;

PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T;  
 PI Pelet C, Schwöpe I, Ziebarth H;  
 DR WPI; 2003-018942/01.  
 XX  
 PT Detecting and differentiating between hematopoietic cell proliferative  
 PT disorders, comprises contacting a target nucleic acid with a reagent  
 PT that distinguishes between methylated and non-methylated CpG  
 PT dinucleotides -  
 PS  
 PS Claim 28; SEQ ID 386; 117pp; English.  
 XX  
 CC The present invention describes a method for detecting and  
 CC differentiating between haematopoietic cell proliferative disorders  
 CC associated with at least 1 gene and/or their regulatory regions in a  
 CC subject. The method comprises contacting a target nucleic acid in a  
 CC biological sample obtained from the subject with at least 1 reagent,  
 CC which distinguishes between methylated and non-methylated CpG  
 CC dinucleotides within the target nucleic acid. AB209861 to AB211118  
 CC represent specifically claimed nucleotide sequences from the present  
 CC invention. Oligonucleotides from the present invention can be used: for  
 CC differentiating between healthy haematopoietic cells and proliferative  
 CC disorder haematopoietic cells; for differentiating between acute  
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for  
 CC determining the cytosine methylation state and/or single nucleotide  
 CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder  
 CC related sequences and their complements; and as primers for the  
 CC amplification of haematopoietic cell proliferation disorder related  
 CC DNA sequences. The nucleotide sequences from the present invention can  
 CC also be used for detecting a predisposition to, differentiation between  
 CC subclasses, diagnosis, prognosis, treatment and/or monitoring of  
 CC haematopoietic cell proliferative disorders. The present method enables  
 CC a highly specific classification of haematopoietic cell proliferative  
 CC disorders allowing for improved and informed treatment of patients.  
 CC  
 SQ Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 other;  
 Query Match 8.7%; Score 54.4; DB 25; Length 8056;  
 Best Local Similarity 46.7%; Pred. No. 0.0043;  
 Matches 172; Conservative 0; Mismatches 196; Indels 0; Gaps 0;  
 QY 162 TTTGAACAATTTCTACTATATTTTACAGAAATCATTCGCTTGAGTTACAGACACA 221  
 Db 6553 TTATTAATAAATTTTATTTATTTATTTAAATAATATATATATATATATATAT 6412  
 QY 222 GTGAGTATATCTTATTTATTTATTTGTCCTCAAGATAGCTATTTATGATGCCAAT 281  
 Db 6413 ATAAATTTATTAATTAATATTTTATTTAAATTAATTTTAAATTTATTTAAAT 6472  
 QY 282 ATTAATTTTATTTGCTTTAGGCGATGATTTCAATTCAGTCAAGTTGCTATATG 341  
 Db 6473 AATAAATAATTTTTTTTATTTTATTTTATTTTAAATTTTGTAAATTTTGTATGAAAA 6532  
 QY 342 GACTACAAAGGTTTAAATTTAAATGAGTAATTTATTAATTAATAAAGGTTGAAGG 401  
 Db 6533 TGTATTTATTTATTTAAATATTTATTTATTTAAATTAATTAATAAAGGAAAA 6592  
 QY 402 AATGTTTGTCTTCAACAACACTGTTGACTTTAGCTTATTTAGTTCAAGTAATGT 461  
 Db 6593 ATTAATTTATTTTATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTA 6652  
 QY 462 TATAGATATTTGCTCATATATGATTAATAAAGAGAGTTTCAAACTTTTACCTATA 521  
 Db 6653 TTTTAAATTTTAAATTAATTTAAATAAATAAATTTTATTTATTTATTTATTTAT 6712  
 QY 522 TTATGCAA 529  
 Db 6713 ATTTATTA 6720

RESULT 5  
 AB210188  
 ID AB210188 standard; DNA; 1501 BP.

XX  
 AC AB210188;  
 XX  
 DT 16-JAN-2003 (first entry)  
 XX  
 DE Haematopoietic cell proliferation disorder related DNA sequence #328.  
 XX  
 KW Human; haematopoietic cell proliferation disorder; cytostatic;  
 KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;  
 KW cytosine methylation state; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 EN WO200277272-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 26-MAR-2002; 2002WO-EP03401.  
 XX  
 PR 26-MAR-2001; 2001US-278333P.  
 PA (EPIG-) EPIGENOMICS AG.  
 PI Berlin K, Braum A, Distler J, Guetig D, Howe A, Mueller J;  
 PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E,  
 PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T;  
 PI Pelet C, Schwöpe I, Ziebarth H;  
 XX  
 XX WPI; 2003-018942/01.  
 DR  
 XX  
 PT Detecting and differentiating between hematopoietic cell proliferative  
 PT disorders, comprises contacting a target nucleic acid with a reagent  
 PT that distinguishes between methylated and non-methylated CpG  
 PT dinucleotides -  
 PS  
 PS Claim 28; SEQ ID 328; 117pp; English.  
 XX  
 CC The present invention describes a method for detecting and  
 CC differentiating between haematopoietic cell proliferative disorders  
 CC associated with at least 1 gene and/or their regulatory regions in a  
 CC subject. The method comprises contacting a target nucleic acid in a  
 CC biological sample obtained from the subject with at least 1 reagent,  
 CC which distinguishes between methylated and non-methylated CpG  
 CC dinucleotides within the target nucleic acid. AB209861 to AB211118  
 CC represent specifically claimed nucleotide sequences from the present  
 CC invention. Oligonucleotides from the present invention can be used: for  
 CC differentiating between healthy haematopoietic cells and proliferative  
 CC disorder haematopoietic cells; for differentiating between acute  
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for  
 CC determining the cytosine methylation state and/or single nucleotide  
 CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder  
 CC related sequences and their complements; and as primers for the  
 CC amplification of haematopoietic cell proliferation disorder related  
 CC DNA sequences. The nucleotide sequences from the present invention can  
 CC also be used for detecting a predisposition to, differentiation between  
 CC subclasses, diagnosis, prognosis, treatment and/or monitoring of  
 CC haematopoietic cell proliferative disorders. The present method enables  
 CC a highly specific classification of haematopoietic cell proliferative  
 CC disorders allowing for improved and informed treatment of patients.  
 CC  
 SQ Sequence 1501 BP; 604 A; 0 C; 130 G; 767 T; 0 other;  
 Query Match 8.6%; Score 53.6; DB 25; Length 1501;  
 Best Local Similarity 47.7%; Pred. No. 0.0047;  
 Matches 187; Conservative 0; Mismatches 204; Indels 1; Gaps 1;  
 QY 131 TTGCTATTTATTCGCGTATGCGATTCATTTTGAACAATTTCTACTATATTTT 190  
 Db 669 TTTTAAATTAATTTATTTATTTAAATAAATGTAATAAATAAATAATTAATATGATGATGA 728  
 QY 191 CAGATCATTTGCTTTAGGTTACAGACAGAGTCAGATATGCTTTATTTATTTATGTTG 250  
 Db 729 TTGTTAATGATATTAATGATTTAAATTAATGATATTTTATTTATTTATTTATTTAA 788

QY 251 CTCAAGATAGCTATTATTATGATGCCAATATCTTTTATGCTGTTACGGCAGTGA 310  
 Db 789 AATTAATAATATTTTGTGTTTATGTTATGTTATTTTGTGATGTTATTTGTTAA 848  
 QY 311 TTTTATTACGCTGCAAGCTGTCATTAAGCACTCAAGAGTTTAAATTAATGA 370  
 Db 849 TTTTAATAATGTTTAAATGTTGAAAAATTAATAAATAATGTTATTTATGATT 908  
 QY 371 GTAAATTAATATAAAGGCTGAAGAAATGTTCTTCAACAAACACTGTTTC 430  
 Db 909 TTTAATTTAATAT-TAAATTAATTTGATGTTGTTAATTTGTTGTTT 967  
 QY 431 GACTTTACGTAAGTTAGTTCAGTAATGTTATAGTAATGTTCCATATATGATTAA 490  
 Db 968 TAAATTTAATATTTAATTTAATTAATTAATGAATAATTTGAAATTTAAATTTG 1027  
 QY 491 AAGGAGAGTTTCAAACTTTTACCATTATAT 522  
 Db 1028 AAATGTAATTTTAAATTTTAAATTTTATTT 1059

## RESULT 6

AAK14127/c  
 ID AAK14127 standard; DNA; 945 BP.

AAK14127;

31-MAR-1999 (first entry)

H. pylori GHPO 243 gene.

GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;

peptic ulcer disease; ss.

Helicobacter pylori.

Key Location/Qualifiers

FT CDS 61..640

FT /\*\*tag= a

W09843478-A1.

08-OCT-1998.

01-APR-1998; 98MO-US06371.

29-JUL-1997; 97US-0902615.

01-APR-1997; 97US-0833457.

24-JUN-1997; 97US-0881227.

(HUMA-) HUMAN GENOME SCI INC.

(INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

Al-Garawi A, Kleantchous H, Miller C, Oomen RP, Tomb J;

WPI; 1998-542293/46.

P-PSDB; AAM98408.

Claim 1; Page 679-680; 2054pp; English.

This sequence represents a polynucleotide of the invention. It was isolated from Helicobacter pylori and encodes a H.pylori GHPO protein. The polypeptides can be used for preventing or treating Helicobacter infections, including acute, chronic, and atrophic gastritis, and peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used for the production of antibodies. The products can also be used for detection and diagnosis.

XX SQ Sequence 945 BP; 298 A; 186 C; 169 G; 292 T; 0 other;

Query Match 8.5%; Score 53; DB 19; Length 945;

Best Local Similarity 46.9%; Pred. No. 0.0059;

Matches 201; Conservative 0; Mismatches 225; Indels 3; Gaps 1;

QY 19 AAAACAGAGAAGCAACCCGGAAGACGTCAGAGAGCTGTTCTGAAGGAGTGCCT 78  
 Db 845 AAAACGAACTCCCTAGCGGAAABAAATCCAAAAGCCAGAGAGAGGCAATGCTCT 786  
 QY 79 AAATCAGAGAGGTTACTAAAGCAATGACTACTGACAGCAGAGAGTCTGGCTGATT 138  
 Db 785 AAGAGCATGAGAGTGGGGGTTTGGGGTTATGGCCGGCTAATTAGATTTTGT 726  
 QY 139 TATTCAGGCGTAATGGAGCTCATTTTGAACAATTTTCTACTATATTTTACAGATCA 198  
 Db 725 TTTTATATATATGATGGGATGAGTGGCTTTAGCGAAATGATGCCATGTGTGAAGA--T 669  
 QY 199 TTTGCGTTGAGGTTACAGCAGATGATATGCTTATTTATATGTTGCTCAGAG 258  
 Db 668 TTTTCCCTAGATTTCACTAAGAAAGGCTTCAAGAGCTGTTTAACTAGCTTAAAGAC 609  
 QY 259 ATAGCTATTTTATGATGCAATATTAATTTTATGCTGTTACGGCAGATTTCAATTA 318  
 Db 608 ACTTTTATTTGCTTTTACGATTTTATCATTTTATGATGTTGTTGCTTTTATCTAAT 549  
 QY 319 CGGTACAGATGTTGCTGATTAATGACTACCAAGCTTTTAAATTAATGAGATTAATT 378  
 Db 548 GTCTTGCAATTTGGCTGGCTCTTTGCCCCCTAAAGTCAATGAGCTTAAATTTTCTAAATC 489  
 QY 379 AATATAATAAAGGGTTGAAGAGATGTTGTTCTCAACAAACACTGTTGCACTTTTA 438  
 Db 488 AACCTATCATATGCGCTCAAAAACCTTTTCTTTTAAAAAAGCTCCTTGATGGAGTTTG 429  
 QY 439 CGTATTTTA 447  
 Db 428 ATCACCCTTA 420

## RESULT 7

ABX6125  
 ID ABX6125 standard; DNA; 1024 BP.

ABX6125;

07-MAY-2003 (first entry)

Helicobacter pylori selected interacting domain (SID) DNA #724.

Protein-protein interaction; ulcer; selected interacting domain;

SID; gene; ds.

Helicobacter pylori.

W020026501-A2.

29-AUG-2002.

28-DEC-2001; 2001MO-BP15428.

02-JAN-2001; 2001US-259302P.

(HYBR-) HYBRIGENICS.

(INSP) INST PASTEUR.

Legrain P, Rain J, Colland F, De Reuse H, Labigne A;

WPI; 2002-674910/72.

P-PSDB; ABUS1381.

New complexes of protein-protein interactions in Helicobacter pylori, useful for identifying modulating compounds for treating or preventing

PT ulcers in mammals -

PS Claim 7, Page 261; 642pp; English.

CC The invention describes a complex of protein-protein interactions in  
CC Helicobacter pylori selected from 421 complexes given in the  
CC specification. The complex of protein-protein interactions are useful  
CC for screening for agents which modulate the interaction of proteins.  
CC Modulating compounds which binds to a targeted bacterial protein may be  
CC used for treating or preventing ulcers in a human or animal. This  
CC sequence encodes a selected interacting domain (SID), identified via  
CC protein-protein interactions.

SO Sequence 1024 BP; 312 A; 187 C; 203 G; 322 T; 0 other;

Query Match 8.5%; Score 53; DB 24; Length 1024;  
Best Local Similarity 46.9%; Pred. No. 0.006;  
Matches 201; Conservative 0; Mismatches 225; Indels 3; Gaps 1;

19 AAAACAGAGAAAGCAACCCGAAAGCAGCTGAGAGAGCTGCTTGAAGGAGTCCCT 78  
141 AAAACGGAAGCTCCCTAGCGCGAATAAATCCAAAAGCCAGAGAAAGGCAATGTGCT 200  
79 AAATCAGAGAGGTTACTTAAGCATCTAGCAGAGAGGATGCTGGGCTTCTATT 138  
201 AAGACATGAGAGTGTGGGGGTTTGGGGTATTGGCCGGCTAATTAGTATTTTGT 260  
139 TATTCAGGCGTAAAGGAGCGTCAATTTTGAACAATTTCTACTATATTTTACAGATCA 198  
261 TTTTATATGTTGGTGGTGGATGGCTTTCAGAAATGTAATCGCATGTGTGAAGA--T 317  
139 TTTGCGTTGAGTTACAGCAGCAGATATGCTTATATTTATTTGCTGCTAAG 258  
318 TTTTCCCTAGATTAGTAAAGAGCGTCAAGCTGTTTAACTGCTAAAGAC 377  
259 ATAGCTATTTTATGATGCAATATTAATTTTATGCTGTTACGCGATGATTCATTA 318  
378 ACTTTTATGCTTTTACCGATTTTAAATCATTTTATGAGTGTGGCTTTTATCTAT 437  
319 CGTGACAGAGTGTGCAATATGACATCAAAAGGTTTAAATTTAAATGAGTAAATT 378  
438 GTCTTGCAATTTGGCTGGCTCTTTCCTTAAAGCATTTGAGCCCTAAATTTCTTAAATC 497  
379 AATATATATAAAGGTTGAAAGGAGATGTTGCTTCTCAACAACAAGCTTGTGCACTTTTA 438  
498 AACCTATCAATGGGTGCAAAAACCTTTTCTTTTAAAAAAGCTCTTGATGGAGTTTG 557  
439 CGTAGTTTA 447  
558 ATCACCCTTA 566

RESULT 8

AAT67941

ID AAT67941 standard; DNA; 1086 BP.

XX AAT67941;

DT 14-JUN-1997 (first entry)

XX H. pylori flagella-associated protein ORF 04gell173orf5.

KM Flagellum; vaccine; prevention; treatment; infection; identification;  
KM binding compound; bacterium; life cycle; activator; bacterial inhibitor;  
KM duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.

XX Helicobacter pylori.

XX Key Location/Qualifiers

FT CDS 1..1086

FT /\*tag= a

FT /note= "no stop codon given"

PM W09640893-A1.

XX 19-DEC-1996.

XX 06-JUN-1996; 96WO-US09122.

XX 01-APR-1996; 96US-0830405.

XX 07-JUN-1995; 95US-0487032.

XX (ASTR ) ASTRA AB.

PI Berglindh OT, Smith D, Mellgaard BL;

DR WPI: 1997-052306/05.

DR P-PSDB; AAM20688.

PT Helicobacter pylori nucleic acid sequences and related  
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
PT infection, and to detect Helicobacter

XX Claim 1; Pages 801-802; 1481pp; English.

XX The present sequence encodes a Helicobacter pylori flagella-  
XX associated protein.

XX The protein may be used in a vaccine to prevent or treat H. pylori  
XX infection or to identify H. pylori polypeptide binding compounds,  
XX useful as potential H. pylori life cycle activators or inhibitors.

XX The genomic sequence of H. pylori (ATCC 55679) was determined from  
XX overlapping contigs generated by mechanically shearing the bacterial  
XX DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
XX and the predicted coding regions defined by computer evaluation. To  
XX identify likely H. pylori antigens for vaccine development, the amino  
XX acid sequences predicted from various ORF were analysed for significant  
XX homology to other known or exported membrane proteins. Having identified  
XX and determined the sequences of interest, particular regions can be  
XX isolated from H. pylori by PCR amplification for recombinant polypeptide  
XX production, e.g. in E. coli hosts.

SO Sequence 1086 BP; 335 A; 193 C; 229 G; 329 T; 0 other;

Query Match 8.5%; Score 53; DB 18; Length 1086;  
Best Local Similarity 46.9%; Pred. No. 0.006;  
Matches 201; Conservative 0; Mismatches 225; Indels 3; Gaps 1;

19 AAAACAGAGAAAGCAACCCGAAAGCAGCTGAGAGAGCTGCTTGAAGGAGTCCCT 78  
28 AAGACGGAAGCTCCCTAGCGCGAATAAATCCAAAAGCCAGAGAAAGGCAATGTGCT 87  
79 AAATCAGAGAGGTTACTTAAGCATCTAGCAGAGAGGATGCTGGGCTTGTATT 138  
88 AAAAGCATGAGAGTGTGGGGTTTATGAGTTATGCTGGCTAATGAGATTTTGT 147  
139 TATTCAGGCGTAAAGGAGCGTCAATTTTGAACAATTTTCTACTATATTTTACAGATCA 198  
148 TTTTATATGAGTGGTGGATGCTTTAGCGATGATGATGCCATGTGTGAAGAATTC 207  
199 TTTGCGTTGAGTTACAGACAGTCAATATCTTATTTATTTATTTATTTGTTGCTCAAG 258  
208 TCCC--TAGATTTTACAAAAGAGCGTTCAAGAGCTTTTAAACAGCTGGCTTAAGAC 264  
259 ATAGCTATTTTATGAGCAATATTAATTTTATTTTATGCTGTTAGGAGATGATTCATTA 318  
265 ACTTTTATTTATGCTTTTCCGCTTTTATATCATTTTATATGAGTGTGGGTTTGTCTAAT 324  
319 CGTGACAGTGTGTCATTAAGCAAGCTTTTAAATTTTAAATGAGTAAATT 378  
325 GTCTTGCAATTTGGCTGGCTCTTTCCTTAAAGCATTTGAGCCCTAAATTTCTTAAATC 384  
379 AATATATATAAAGGTTGAAAGGAGATGTTGCTTCTCAACAACAAGCTTGTGCACTTTTA 438  
385 AACCTATCAATGGGTGCAAAAACCTTTTCTTTTAAAAAAGATCTTGATGGAGTTTG 444  
439 CGTAGTTTA 447



CC differences serving as basis for diagnosis and/or prognosis events which  
 CC are disadvantageous to patients. The present sequence is one of the  
 CC 533 genomic sequences derived from tumour suppressor genes and  
 CC oncogenes. Sequences with even numbered Seq ID numbers are the  
 CC complementary sequence of the corresponding odd numbered sequence (e.g.  
 CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence  
 CC is missing).  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 5297 BP; 1353 A; 107 C; 1126 G; 2711 T; 0 other;

Query Match 8.4%; Score 52; DB 22; Length 5297;  
 Best Local Similarity 46.6%; Pred. No. 0.014;  
 Matches 203; Conservative 0; Mismatches 230; Indels 3; Gaps 1;

QY 158 GTCATTGGAACATTTCTACTATTTTACAGAACATTCATTGGTTGAGTTACAG 217  
 Db 2669 GTAATTTTGGATTTGGTTTGGTTTATTTATTTATTTAGTTTATTTAGGT 2728  
 QY 218 CACAGTCAGTATA--TGCTTATTTATTTATTTATTTATTTATTTATTTATTTA 274  
 Db 2729 TATATTGGAAATAGAGATTTAGTTTGTATTTTGTATTTTGAAGTATTTTGA 2788  
 QY 275 TGCATATTTACTTTTATTTGCTGTACGCGATGATTTCAATTCAGTTGAGT 334  
 Db 2789 GATGTTTATTTATTTTATTTTATTTATTTATTTATTTATTTTATTTGATG 2848  
 QY 335 CATTTAGCACTACAAAGTTTCTAATTTAATTTAATTTAATTTAATTTAATTTA 394  
 Db 2849 AAAATTAATTTTGTTTTGTATTTTATTTATTTATTTATTTATTTATTTATTT 2908  
 QY 395 TGAAGGATTTGCTTCTCAACAACAACCTTTGCACTTTTACGTTAGTTTACG 454  
 Db 2909 TTAATTTAGAAATAGTTTATTTAGTATTTAGTATTTAGTATTTAGTATTTA 2968  
 QY 455 TAAATGTTATAGGATTTGTTCCATTTATGATTTAATAAGAGAGATTTCAAC 514  
 Db 2969 TTGAATTTGTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 3028  
 QY 515 CATTTATTTATTTAGCAAGTCTTCAGGTTGCGAGATTTATTTGTTATTTAC 574  
 Db 3029 GATTTAGGTTAATTTGTTATTTATTTATTTATTTATTTATTTATTTAT 3088  
 QY 575 TTGTTTATTTATAGCT 590  
 Db 3089 ATTTTATTTTATAGGT 3104

RESULT 11  
 ABL33559  
 ID ABL33559 standard; DNA; 7128 BP.

XX ABL33559;

XX 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 1532.

XX Human, immune system disease; cytosine methylation; antiaesthetic;  
 XX antiarteriosclerotic; anti-HIV; anticonvulsant; ophthalmological;  
 XX antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 XX gene; ds.

XX Homo sapiens.

OS WO000200928-A2.  
 XX  
 PN

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPICENOMICS AG.

XX Olek A. Piepenbrock C. Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful  
 XX for diagnosis and treatment of diseases associated with abnormal  
 XX cytosine methylation -

XX Claim 1; SEQ ID NO 1532; 32pp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.

XX Sequence 7128 BP; 1875 A; 207 C; 1610 G; 3436 T; 0 other;

Query Match 8.4%; Score 52; DB 24; Length 7128;  
 Best Local Similarity 46.5%; Pred. No. 0.015;  
 Matches 240; Conservative 0; Mismatches 270; Indels 6; Gaps 2;

QY 78 TAAATCAGAAAGGTTCTAAGGATTTGACTGACAGAGGATGCTGGGCTTGTAT 137  
 Db 4809 TAAATTTATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 4868  
 QY 138 TTAATTCAGGCTAATGAGAGCTCATTTTGAACAATTTTCTACATATTTTACAG 197  
 Db 4869 TTAGTATGATGATGATGATTTATTTATTTATTTATTTATTTATTTATTTA 4928  
 QY 198 ATTTCGTTGAGGTTACAGACAGTCAGTATGCTTATTTATTTATTTATTTAT 257  
 Db 4929 TTTAATATGATTTAATAATTAAGTTATTAATGATTTATTTATTTATTTAT 4988  
 QY 258 GATAGCTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 317  
 Db 4989 TAT-----TATGATGAGAAATTTATTTATTTATTTATTTATTTATTTAT 5043  
 QY 318 ACGTGTACAGTTGGTGTATTTATTTATTTATTTATTTATTTATTTATTTAT 377  
 Db 5044 TAGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 5102  
 QY 378 TAAATTTAATAAGGTTGAAGAAGATTTGCTTCAACAACAACCTTTGCACTTT 437  
 Db 5103 TTAGATTTGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 5162  
 QY 438 ACGTATTTAGTTCAAGTATTTATTTATTTATTTATTTATTTATTTATTTAT 497  
 Db 5163 GGGGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 5222  
 QY 498 GTTTTCAACCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 557  
 Db 5223 GATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 5282  
 QY 558 TAAATCAGAAATGATTTGTTTATTTATTTATTTATTTATTTATTTATTTAT 593  
 Db 5283 TAAATTTAATAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 5318

RESULT 12

AB067150  
 ID AB067150 standard; DNA; 40324 BP.  
 AC AB067150;  
 XX  
 XX 28-AUG-2002 (first entry)  
 DE Human angiogenesis associated polynucleotide SEQ ID NO 180.  
 XX  
 XX Human; angiogenesis; methylation; eye disease; glaucoma; tumour;  
 KM inflammation; rheumatoid arthritis; diabetic retinopathy; antileucoderis;  
 KM macular degeneration; inflammatory bowel disease; Crohn's disease;  
 KM antileukemic; antileukemic; antidiabetic; antipsoriatic;  
 KM antileukosclerotic; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200246454-A2.  
 XX  
 PD 13-JUN-2002.  
 XX  
 XX 06-DEC-2001; 2001WO-EP14320.  
 XX  
 PR 06-DEC-2000; 2000DE-1061338.  
 XX  
 XX (EPIC-) EPIDEMIOLOGICS AG.  
 XX  
 PI Schacht O;  
 XX  
 DR WPI; 2002-500450/53.  
 XX  
 XX New nucleic acid fragments from chemically treated  
 PT angiogenesis-associated genes, useful for determining methylation  
 PT status, e.g. in diagnosis or treatment of cancer -  
 XX  
 PS Claim 1; SEQ ID NO 180; 41bp + Sequence Listing; German.  
 XX  
 CC The invention relates to a nucleic acid (I) comprising a segment of 18  
 CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)  
 CC having sequences (AB066971-AB067178) or their complements. (II), also  
 CC related oligomers, are used to evaluate the methylation status and/or  
 CC single-nucleotide polymorphisms, in angiogenesis-related genes, for  
 CC diagnosis and treatment of eye diseases, proliferative retinopathy,  
 CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,  
 CC diabetic retinopathy, macular degeneration caused by neovascularisation,  
 CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and  
 CC Crohn's disease.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SO Sequence 40324 BP; 12266 A; 436 C; 7652 G; 19970 T; 0 other;  
 Query Match 8.4%; Score 52; DB 24; Length 40324;  
 Best Local Similarity 46.5%; Pred. No. 0.022;  
 Matches 240; Conservative 0; Mismatches 270; Indels 6; Gaps 2;  
 QY 78 TAAATCAGAGAGCTTACTAAGATGACATGACGACGAGGATGCTGGGCTTGTAT 137  
 DB 38005 TAAATTTATAGATTATTTAATTTGATTTTAAAGTTAAATGATGATTAATAT 38064  
 QY 138 TTTATCAGGCGTATGAGAGCTCATTTTGAACAATTTTACATATTTTTCAGATC 197  
 DB 38065 TTACTTTAGTATGATGATGATTTATTTTGTATTTATTTTAAATTTTAAATTAAT 38124  
 QY 198 ATTTCGTTGAGGTTACAGACAGCAGTATATGCTTTATTTATTTATTTGTTCTCAGA 257  
 DB 38125 TTTTAAATATGATTTAAATTAAGATTAATATGATTAATGATTTTATTTGATTTGTT 38184  
 QY 258 GATAGCATTTTATGATGCAATATTAATTTTATTTGCTGTACGCGATGATTTTATTT 317  
 DB 38185 TAT-----TATGATGAGAAAATTTTGAAGTTTATTTTATGAGATTTAATTTT 38239

QY 318 ACGTACAGAGTTCGTCATTTATGACTACACAAAGTTTAAATTTAAATGAGTAAT 377  
 DB 38240 TAGTTTATTTATTTAGTATTTATTTATTTATTAATA-TTATTTATGAGTATTTATTTATG 38298  
 QY 378 TAAATTAATTAAGAGGTTGAAAGATGTTTGTCTTCAACAAACACTTGTTCACATTT 437  
 DB 38299 TTAGATTTATTTTATTTATTTAGGATTAATAAGTAAATATATATATGTTTATTTATTTAT 38358  
 QY 438 ACGTATTTATTTCAATTAATTTTATTTAGGATTTGTCATATATGATTTAAAGAGA 497  
 DB 38359 GGGGTTTATTTATTTATTTAGTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 38418  
 QY 498 GTTTCAACTTTTATTTACATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 38478  
 DB 38419 GATTTTATTTGTTTGTGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 38478  
 QY 558 TAAATCAGAAATGATGATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 593  
 DB 38479 TAAATTTTAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 38514  
 RESULT 13  
 ABL32164  
 ID ABL32164 standard; DNA; 5523 BP.  
 AC ABL32164;  
 XX  
 XX 26-MAR-2002 (first entry)  
 XX  
 XX Human immune system associated gene SEQ ID NO: 137.  
 DE  
 XX Human; immune system disease; cytosine methylation; antileukemic;  
 KM antileukosclerotic; antileukemic; cytostatic; neoplastic;  
 KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KM antileukemic; antileukemic; antidiabetic; antipsoriatic;  
 KM antileukemic; cancer; eye disease; arteriosclerosis; anaemia;  
 KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KM gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200200928-A2.  
 XX  
 PD 03-JUN-2002.  
 XX  
 XX 02-JUL-2001; 2001WO-EP07537.  
 XX  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIC-) EPIDEMIOLOGICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 DR WPI; 2002-130909/17.  
 XX  
 XX Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation -  
 XX  
 PS Claim 1; SEQ ID NO 137; 32pp + Sequence Listing; German.  
 XX  
 CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 CC  
 SO Sequence 5523 BP; 1378 A; 119 C; 1366 G; 2660 T; 0 other;



Query Match 8.3%; Score 51.8; DB 24; Length 5523;  
 Best Local Similarity 45.6%; Pred. No. 0.016;  
 Matches 182; Conservative 0; Mismatches 217; Indels 0; Gaps 0;

185 TTTTACGAATCATTTGGGTTGGGTTACAGCACTGCTGATATGCTTATTTATTT 244  
 |||||  
 726 TTTTATATATAATTTTGTGTTGATGTTTGGATTTTGTATATGTTTAAATGTT 785  
 |||||  
 245 ATGTGCTCAAGAGATAGCTATTTTATATGATGCCAATATTACTTTATTTGCTGTTACGG 304  
 |||||  
 786 ATGAGCTTGAAGTTAAATATATTTATGCTTTGTTTGTATTTTACGTTTAAATTTTATAG 845  
 |||||  
 305 CATGATTTTCATTACGCTGACAACTGGTGCTATATGACATCAAAAGTTTAAATTTA 364  
 |||||  
 846 AATTAAGTAGTTAAGTTGAATGGGTTTAAAGAGATCTGTTATTTTGTGTTTATATAT 905  
 |||||  
 365 AATGAGATTAATTTAATTAATAAAGGTTGAAGGATGTTTCTCTCAACAAAC 424  
 |||||  
 906 ATAGATGAGGATTTGTTATTTAGAGTCTGAGAAATGTTTATGCTTTTATGAT 965  
 |||||  
 425 TTGTCGACTTTTACGTTTATGTTTCAAGTATTTGTAAGATTTGTTCCATATATGA 484  
 |||||  
 966 TTTTATTTTATTTTATAGTTTAACTTATCTGATTTTAAATGAAGATGTAATATATA 1025  
 |||||  
 485 TTATATAAGAGAGATTTTCAAACTTTTACCATTAATATATGCAAGTCTTCAGGTGG 544  
 |||||  
 1026 GTAGTAATATTTAAAGATTAAGATGATGATGATATGCAAGAAATTTATATATGAT 1085  
 |||||  
 545 CAGATTATATGCTTAAATACAGAAATGATCTGTTTAT 583  
 |||||  
 1086 TATATTAATTAATTAATTAATGATGTTATGTTGTTGATAT 1124  
 |||||

RESULT 14  
 ABL92318  
 ID ABL92318 standard; DNA; 73334 BP.

XX ABL92318;  
 AC  
 XX  
 DT 01-JUL-2002 (first entry)  
 XX  
 DE Chemically treated DNA repair gene fragment#64.  
 XX  
 XX DNA repair; cytosine methylation; PMS2L1; PMS2L12; PMS2L3;  
 KM PMS2; L4; PMS2L5; PMS2L6; MGMT; MSH2; NUDT1; TDG; INPPL1; RFC4;  
 KM DDIR1L; FANCB; XRCC8; ataxia telangiectasia; aging; Bloom's syndrome;  
 KM Cockayne syndrome; Nijmegen breakage syndrome; Werner syndrome;  
 KM immunodeficiency; trichiodystrophy; Fanconi's anaemia; solid tumour;  
 cancer; ds.  
 KM  
 XX  
 OS Unidentified.  
 XX  
 PN WO200181622-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PD 06-APR-2001; 2001WO-EP03972.  
 PF  
 XX 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 XX (EPIC-) EPIGENOMICS AG.  
 PA  
 XX Olek A, Piepenbrock C, Berlin K;  
 PI  
 XX WPI; 2002-034446/04.  
 DR  
 XX  
 XX New nucleic acid derived from genes associated with DNA repair, useful  
 PT for diagnosis, e.g. of ataxia telangiectasia, by determination of  
 PT cytosine methylation -

XX Claim 1; SEQ ID NO 127; 25pp + sequence listing; English.  
 PS  
 XX  
 CC The invention relates to nucleic acids containing a sequence of at least  
 CC 18 nucleotides of chemically treated DNA of genes associated with DNA  
 CC repair, and their complements. The invention also relates to nucleic  
 CC acids comprising at least 18 base pairs of the chemically pretreated DNA  
 CC of genes associated with DNA repair selected from PMS2L1, PMS2L12,  
 CC PMS2L3, PMS2, L4, PMS2L5, PMS2L6, MGMT, MSH2, NUDT1, TDG, INPPL1,  
 CC RFC4, DDIR1L, FANCB, or XRCC8. Nucleic acids of the invention and related  
 CC oligomers, are useful for diagnosis of diseases associated with gene  
 CC repair, specifically ataxia telangiectasia, aging, Bloom's syndrome,  
 CC Cockayne syndrome, Nijmegen breakage syndrome or Werner syndrome,  
 CC immunodeficiency, trichiodystrophy, Fanconi's anaemia, solid tumours  
 CC and cancer, particularly by determining status of cytosine methylation  
 CC and/or by detecting single-nucleotide polymorphisms. Determination of  
 CC individual methylation patterns may allow development of individualised  
 CC therapies. The sequences given in records ABL92192-ABL92335 represent  
 CC chemically pre-treated DNA fragments from genes associated with DNA  
 CC repair, and their complements.  
 CC Note: The sequence data for this patent is not represented in the  
 CC specification, but is based on sequence information supplied by the  
 CC European Patent Office.  
 XX  
 SQ Sequence 73334 BP; 18968 A; 858 C; 15329 G; 38179 T; 0 other;

Query Match 8.3%; Score 51.8; DB 24; Length 73334;  
 Best Local Similarity 49.8%; Pred. No. 0.028;  
 Matches 131; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

264 TATTTATGATGCAATATTAATTAATTTATGCTGTTACGCAATGATTCATTACGTGT 323  
 |||||  
 57666 TAAATATTAATTAATTAATTAATTAATTTTATGATGATTTTAAATTAATTAATTA 57725  
 |||||  
 324 ACAAGTTGTCATTATGACATCAAGGTTTAAATTAATTAATGAATTAATTAAT 383  
 |||||  
 57726 TAAATATTAATTAATTAATTAATTTTATGATGATTTTAAATTAATTAATTAAT 57785  
 |||||  
 384 AATTAAGGTTGAAGGAATGTTGCTTCACAAACAACTTGTGACCTTTACGAT 443  
 |||||  
 57786 TTTAGTAATTAATTAATTAATTAATGATTTTATTAAGGAAGGAATTTGTTAT 57845  
 |||||  
 444 TTTAGTCAAGTAATTTGTTATGATTTGCTCATATATGATTTAAAGGAGATTTTC 503  
 |||||  
 57846 TTAAGTGAATTTGATTAATTAATTTTATTTATTTATTTATTTAATTAATTAATTA 57905  
 |||||  
 504 AAACCTTTTACCATTAATTAATG 526  
 |||||  
 57906 AAATATTTTGTGATTAAGTTATG 57928  
 |||||

RESULT 15  
 ABL34124  
 ID ABL34124 standard; DNA; 73334 BP.

XX ABL34124;  
 AC  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 2097.  
 XX  
 XX Human; immune system disease; cytosine methylation; antiasthmatic;  
 KM antiarteriosclerotic; antianemic; cytostatic; noctropic;  
 KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KM antirheumatic; antiarthritic; antidiabetic; antipoxiatic;  
 KM antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;  
 KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KM gene; ds.  
 KM  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO200200928-A2.

XX 03-JAN-2002.  
PD  
XX  
PF 02-JUL-2001; 2001WO-EP07537.  
XX  
XX 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPIC-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
DR WPI; 2002-130909/17.  
XX  
XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation -  
XX  
XX  
PS Claim 1; SEQ ID NO 2097; 32pp + Sequence Listing; German.  
XX  
XX The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
SQ Sequence 73334 BP; 18968 A; 858 C; 15329 G; 38179 T; 0 other;  
Query Match 8.3%; Score 51.8; DB 24; Length 73334;  
Best Local Similarity 49.8%; Pred. No. 0.028;  
Matches 133; Conservative 0; Mismatches 132; Indels 0; Gaps 0;  
QY 264 TATTTATGATGCCAATATTAATCTTTTATGCTGTACGCGATTCATTACGTGT 323  
DB 57666 TAAGTATTTAATGGAATTAAGAAATTTTATGATGATTTTAAAAAAGTTA 57725  
QY 324 ACAAGTTGTCATATGACGCTCAAGCTTTTAAATTTAAATGAGTAAATTAAT 383  
DB 57726 TAAATATAGGGGTAATATGATTTATTTATTTTGAAGTAAGTAAGAAATTAAT 57785  
QY 384 AATTAAGGCTGTAAGAAAGATGTTGCTCTCAACAACACTGTTGACCTTTAGCTAG 443  
DB 57786 TTTAGTAAATTAATAAATAGTATTAGCTTTTATTAAGAAAGAAATTTTGTAA 57845  
QY 444 TTTAGTCAAGTATTTGTTATAGTATGTTCCATATATGATTAATAAAGAGATTTTC 503  
DB 57846 TTAAGTGAATGTAATTAATTTTAAATTTTATTTGATTAATTTAGGGGTTT 57905  
QY 504 AAACCTTTTACATTAATATATG 526  
DB 57906 AAATATTTTGTATGATGATATG 57928  
RESULT 16  
ABL33299  
ID ABL33299 standard; DNA; 12025 BP.  
XX  
AC ABL33299;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Human immune system associated gene SEQ ID NO: 1272.  
XX  
XX Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiasthmatic; anti-HIV; anticonvulsant; ophthalmological;  
KW antineoplastic; anti-HIV; anticonvulsant; ophthalmological;  
KW antineoplastic; antiasthmatic; antidiabetic; antipsoriatic;  
KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;

KW gene; ds.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200200928-A2.  
PN  
XX  
XX 03-JAN-2002.  
PD  
XX  
XX 02-JUL-2001; 2001WO-EP07537.  
PF  
XX  
XX 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPIC-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
DR WPI; 2002-130909/17.  
XX  
XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation -  
XX  
XX  
PS Claim 1; SEQ ID NO 1272; 32pp + Sequence Listing; German.  
XX  
XX The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
SQ Sequence 12025 BP; 4009 A; 84 C; 2126 G; 5806 T; 0 other;  
Query Match 8.2%; Score 51.2; DB 24; Length 12025;  
Best Local Similarity 46.5%; Pred. No. 0.026;  
Matches 201; Conservative 0; Mismatches 228; Indels 3; Gaps 1;  
QY 163 TTGAAACAATTTTCTACTATATTTTACAGAAATCTTGGTTGAGGTTACAGACAG 222  
DB 203 TTTTACAGAAATTTTACGTTGATTTTAAATTTTATGATGATGTTGATTA 262  
QY 223 TCAGTATATGCTTATTTATTTATGTTGCTCAAGATAGCTATTTATGATGCCAAT 282  
DB 263 TAGATATTTGGGTTTATTTTATTTTAAAGTTATTTTATTTTAAATTAATGAATA 322  
QY 283 TTAATTTTATTTGCTGTTACGCGATGATTTCAATGCTGTAACAAGTTGGCATTAATG 342  
DB 323 TGAATTTAAGTAGAGATTTATTTAGGAATTTGTTGATTAATTAATGATTAATTCGTA 382  
QY 343 ACTACAAGGTTTAAATTTAAATGAGTAATTTAATATTAATAAAGGTTGAAGGA 402  
DB 383 ATTAAGCTATATTTTAAATGTAAGTAAATGTAAGTAAATTTTAAAGTTAAGG 442  
QY 403 ATGTTGCTTCAACAACAACCTTG--TTCACTTTACGTTAGTCAAGTAAT 459  
DB 443 AAGTTTATTAATTTGATATTTTGTATTTAAGAAATTTTATTTAGGTATTTATTT 502  
QY 460 GTTATAGTATTTGCTTCAATATATGATTAATAAAGAGATTTTCAACTTTTACCATTA 519  
DB 503 TATTAATGTAATTAATTAATTTTATGATTAAGTAATTTTATGCTGTTATATGTA 562  
QY 520 TATTAATGCAAGTCTTCAAGTGTGCGAATTAATGCTTAATACAGAAATGACTTGT 579  
DB 563 TTTTAATTAATGTTGTTGGAATTAATAATATGTTTATTAATGAATTTAGTATAGGT 622  
QY 580 TATATACGCTA 591  
DB 623 TTAAGTTGGTA 634

RESULT 17  
 ABL33031  
 ID ABL33031 standard; DNA; 6118 BP.  
 XX  
 AC ABL33031;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 1004.  
 XX  
 KW Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; anti-anaemic; cytosine; neoplastic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200200928-A2.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 02-JUL-2001; 2001WO-EP07537.  
 XX  
 PR 30-JUN-2000; 2000DE-1032529.  
 XX  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 DR WPI; 2002-130909/17.  
 XX  
 PT Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation -  
 XX  
 PS Claim 1; SEQ ID NO 1004; 32pp + Sequence Listing; German.  
 XX  
 CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 XX  
 SO Sequence 6118 BP; 1751 A; 141 C; 1229 G; 2997 T; 0 other;  
 XX  
 Query Match 8.2%; Score 51; DB 24; Length 6118;  
 Best Local Similarity 50.5%; Pred. No. 0.025;  
 Matches 151; Conservative 0; Mismatches 145; Indels 3; Gaps 1;  
 XX  
 QY 227 TATATGCTTATTTATTTATGCTCAAGAGATGACTTTTATGATGCATTTATAC 286  
 DB |||||  
 DB 4654 TAAATGCTTATTTTATTTTATTTATTTATTTATTTATTTATTTATTTT 4713  
 QY 287 TTTTATTTGCTGTAACGCGATGATTTGATTCATTCGTAAGTGGTGCATTTAGACTA 346  
 DB |||||  
 DB 4714 TTTTATTTGCTGTTTGTTTTAAAT---TTTATTTAAGAGTTATGTTTATTTTA 4770  
 QY 347 CAAAGCTTTTAAATTTAAATGAGTAATTTATATATAAAGGTTGAAGGATGT 406  
 DB |||||  
 DB 4771 GAGATTTTAAATGAGAAATGATAATATATAGGTTAGTGAAGGTTAAGGATTTG 4830  
 QY 407 TTGCTTCAACAACAACCTTTGACCTTTACGAGTTAGTTAGTCAATTAATGTTATAG 466  
 DB |||||  
 DB 4831 AAGTTTCATTTTGGGATTAAGTTTATATATATGTTAGTTAATTTATTTT 4890

QY 467 GTATTTCCATATATGATTTATAAAGAGAGCTTTCAACCTTTTACCATTTATTTAT 525  
 DB |||||  
 DB 4891 TTAATGTTTATTTTGAATAAATATGAAATATATATATATATATATATATAGTCT 4949  
 XX  
 AC ABL33971;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 1944.  
 XX  
 KW Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; anti-anaemic; cytosine; neoplastic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200200928-A2.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 02-JUL-2001; 2001WO-EP07537.  
 XX  
 PR 30-JUN-2000; 2000DE-1032529.  
 XX  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 DR WPI; 2002-130909/17.  
 XX  
 PT Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation -  
 XX  
 PS Claim 1; SEQ ID NO 1944; 32pp + Sequence Listing; German.  
 XX  
 CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 XX  
 SO Sequence 7921 BP; 2387 A; 83 C; 1706 G; 3745 T; 0 other;  
 XX  
 Query Match 8.2%; Score 51; DB 24; Length 7921;  
 Best Local Similarity 46.5%; Pred. No. 0.027;  
 Matches 165; Conservative 0; Mismatches 190; Indels 0; Gaps 0;  
 XX  
 QY 159 TCAATTTGAACAATTTTCTACTATATTTTATACAGATCATTTGCTTGAAGTTACAGC 218  
 DB |||||  
 DB 2153 TGAATTTTAAACGTAAGTTTATGATTTGTAAGTAATTTTGTATTTTATTTTATGAGGT 2212  
 QY 219 ACAAGTATATGCTTTATTTATTTATGCTCAAGATAGCTATTTATTTATGATGCC 278  
 DB |||||  
 DB 2213 TGGGGGGGATGAGTTAATTTATGTAAGTTTATATAATATATATGTTATGTT 2272  
 QY 279 AATTTACTTTTATTTGCTTTACGCGATGATTTTCAATTCATTCGTTACAGTTGGTGCA 338  
 DB |||||  
 DB 2273 AGTATATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTT 2332



XX DR WPI; 2002-130909/17.  
 XX PT Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation  
 XX PS Claim 1, SEQ ID NO 1804; 32bp + Sequence Listing; German.  
 XX CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/nutritional bowel  
 CC diseases. The present sequence is a gene of the invention.  
 XX SQ Sequence 6126 BP; 1772 A; 88 C; 1318 G; 2948 T; 0 other;  
 Query Match 8.1%; Score 50.6; DB 24; Length 6126;  
 Best Local Similarity 46.7%; Pred.No. 0.031;  
 Matches 161; Conservative 0; Mismatches 184; Indels 0; Gaps 0;  
 QY 203 GGTGAGGTTACACACAGTCAGTATGCTTTATTTATTTATGCTCAGAGATAG 262  
 DB 70 GGTGAGGTTAAATTTATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 129  
 QY 263 CTATTTTTCATGTCGCAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 322  
 DB 130 TTTTATTTGTTTGTATTTAAAGTTTGAATTTATTTATTTATTTATTTATTTATTTATTT 189  
 QY 323 TACAAGTTGTCATATGATGACCTAACAAGTTTAAATTTAAATTTAAATTTAAATTTAAATTT 382  
 DB 190 ATAGAGAAATTAAGTTATTTGTTATGAAAAATTTATTTGTTAAATTTATTTATTTATTT 249.  
 QY 383 TAATTAAGGTTGAAGAAATGTTGCTTCACAACAACCTGTTGACCTTTACGTA 442  
 DB 250 AAAATTTAAGTTTAAAAAAGTTAAGTTATGAGATTTTATTTTATTTTAAATTAAT 309  
 QY 443 GTTTAGTTCAAGTAATTTGTTATGATTTGTTCCATATATGATTTAAAGAGAGTTT 502  
 DB 310 TTTAAATTAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 369  
 QY 503 CAACCTTTTACATTAATTAATGCAAGTCCCTCAGGTCGAG 547  
 DB 370 ATAAATATTTGTTATTAATTTTATTAATTAATTTGAAGTTGAGAG 414  
 RESULT 21  
 AAS45480  
 ID AAS45480 standard; DNA; 11047 BP.  
 XX AC AAS45480;  
 XX DT 18-DEC-2001 (first entry)  
 XX DE Chemically pretreated genomic DNA associated with cell cycle #93.  
 XX KW Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;  
 KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;  
 KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;  
 KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritis;  
 KW immunosuppressive; antitumour; cytosolic; antiarteriosclerotic; ds;  
 XX PCR primer.  
 XX OS Homo sapiens.  
 XX PN WO200168911-A2.  
 XX PD 20-SEP-2001.  
 XX PF 15-MAR-2001; 2001WO-EP02945.

XX PR 15-MAR-2000; 2000DE-1013847.  
 PR 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX PA (EPIC-) EPIGENOMICS AG.  
 XX PI Olek A, Piepenbrock C, Berlin K;  
 XX DR WPI; 2001-602751/68.  
 XX PT Designing primers and probes for analysing diseases associated with  
 PT cytosine methylation state e.g. arthritis, cancer, aging,  
 PT arteriosclerosis comprising fragments of chemically modified genes.  
 PT associated with cell cycle  
 XX PS Claim 1, SEQ ID No 185; 28bp; English.  
 XX CC Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA  
 CC molecules associated with the cell cycle and specific PCR primers of the  
 CC invention. The sequences are useful for detecting the methylation state  
 CC of all CpG dinucleotides in a sequence and therefore for analysing  
 CC associated diseases. By analysing cytosine methylations in the pretreated  
 CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy  
 CC of existing diseases or the predisposition to specific diseases can be  
 CC ascertained. The parameters may be compared to another set of genetic  
 CC and/or epigenetic parameters, the differences serving as basis for  
 CC diagnosis and/or prognosis events which are disadvantageous to patients.  
 CC The sequences of the invention are useful for the diagnosis and therapy  
 CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,  
 CC aging, glomerular disease, Lewy body disease, arthritis,  
 CC arteriosclerosis, solid tumours and cancers.  
 XX SQ Sequence 11047 BP; 3237 A; 248 C; 2358 G; 5204 T; 0 other;  
 Query Match 8.1%; Score 50.6; DB 22; Length 11047;  
 Best Local Similarity 46.9%; Pred.No. 0.036;  
 Matches 158; Conservative 0; Mismatches 179; Indels 0; Gaps 0;  
 QY 164 TTGAACAATTTCTACTATTTATTTTACAGAAATCATTTGGTTGAGTTACGACAGT 223  
 DB 10382 TTGAATGAATTTAGAGATTAATTTTGAATTAATTTATTTGTTAGTTGTTAGAGA 10441  
 QY 224 CAGTATATGCTTATTTATTTATTTATGTTGCTCAGAGATAGCTATTTATTTGCAATAT 283  
 DB 10442 GCGATTAATGTTATTTAGTTATTTATTTGTTGAATTTTATTTGTTGAAGTTTG 10501  
 QY 284 TACTTTTATTTGCTGTACGGCATGATTTCTATGACAGAAATTTGCGATTTATNGA 343  
 DB 10502 TTTTATTTTAAAGTAATGTAATTAAGAAATTTAGATTAATTTGTTTATTTT 10561  
 QY 344 CTACAAGGTTTAAATTTAAATGAGTAATTTAATTAATTAAGGTTGAAAGAA 403  
 DB 10562 TTGATTAATTTTATTTTAAATTTATTTATTTATTTGCGTTAAGAAAGTTGAAATTA 10621  
 QY 404 TGTTCCTTCTCAACAACACTGTTGCACTTTAGCTGATTTAGTTCAAGTAATTTGTTA 463  
 DB 10622 TTTTATTTTAAATTTGTAATTTGTTTGTGTTAGTGAAGGTAATTTAGTAAGGATTTGTTT 10681  
 QY 464 TAGGATTTGTTCAATATGATTAATAAGAGAGATT 500  
 DB 10682 TTGATGTAATTTTATTTGATTTTATTTAGTTATTTGGAATTT 10718  
 RESULT 22  
 ABL33985  
 ID ABL33985 standard; DNA; 11047 BP.  
 XX AC ABL33985;  
 XX DT 26-MAR-2002 (first entry)

XX DE Human immune system associated gene SEQ ID NO: 1958.  
 XX XX  
 KW Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antianemic; cytosstatic; noctropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; optalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antitumour; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.  
 XX OS Homo sapiens.  
 XX XX  
 PN MO200200928-A2.  
 XX XX  
 PD 03-JAN-2002.  
 XX XX  
 PF 02-JUL-2001; 2001WO-EP07537.  
 XX XX  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX XX  
 PA (EPig-) EPIGENOMICS AG.  
 XX XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 DR WPI; 2002-130909/17.  
 XX XX  
 PT Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation -  
 XX XX  
 PS Claim 1; SEQ ID NO 1958; 32pp + Sequence Listing; German.  
 XX XX  
 CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 CC XX  
 SQ Sequence 11047 BP; 3237 A; 248 C; 2358 G; 5204 T; 0 other;  
 Query Match 8.1%; Score 50.6; DB 24; Length 11047;  
 Best Local Similarity 46.9%; Pred. No. 0.036;  
 Matches 158; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

QY 164 TTGAACAATTTCTACTATATTTTACAGATCATTTCGTTGAGGTACACACACT 223  
 DB 10382 TTGAAGAATAATAGATTAATATTTTGAATAATATATATTTTGAATTTGTTAGTAGAGA 10441  
 QY 224 CAGTAATGCTTAATTTATTTATTTGCTCAAGATAGCTAATTTTATGATGCATAT 283  
 DB 10442 GCGATTAATTTATTTATTTGTTATTTGTTAGATTTTATTTTGTGTTGAAGTTGG 10501  
 QY 284 TACTTTTATTTGCTTTAGCGCATGATTTCAATTCAGTGAACAAGTTGTCATTAATGA 343  
 DB 10502 TTTTATTTTAAAGTAATGTAAGTAAAGATTAAGATTAAGTGAATGTTGGTTTTTT 10561  
 QY 344 CTACAAAGTTTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 403  
 DB 10562 TTGATTAATTTTATTTTATTTTAAATTTATTTTGGTTAAAGAAAGTTTAAATTATTA 10621  
 QY 404 TGTTCCTTCACAAACAACCTGTCGACTTTTACAGTTAGTTCAAGTAATTTGTA 463  
 DB 10622 TTTTATTTTATTTATTTATTTGTTTGTAGTAGTAGTAATTAATTAATTAATTAATTA 10681  
 QY 464 TAGGTATTTGTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 500  
 DB 10682 TTGATTAATTTTATTTTATTTTATTTTATTTTGGTAATTT 10718

RESULT 23  
 ID ABRK28402 standard; DNA; 11047 BP.  
 XX ABRK28402;  
 AC ABRK28402;  
 XX XX  
 DT 23-APR-2002 (first entry)  
 XX XX  
 DE DNA transcription associated complementary genomic DNA #138.  
 XX XX  
 KW DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;  
 KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;  
 KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;  
 KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;  
 KW immunological disorder; Werner syndrome; developmental disorder;  
 KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;  
 KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;  
 KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;  
 KW angiogenesis; congenital heart disease; HDR syndrome; gene therapy;  
 KW polyuramine disorder; solid tumour.  
 XX XX  
 OS Unidentified.  
 XX XX  
 PN MO200192565-A2.  
 XX XX  
 PD 06-DEC-2001.  
 XX XX  
 PF 06-APR-2001; 2001WO-EP03973.  
 XX XX  
 PR 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX XX  
 PA (EPig-) EPIGENOMICS AG.  
 XX XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 DR WPI; 2002-090046/12.  
 XX XX  
 PT New nucleic acids or oligomers, useful for diagnosing or treating  
 PT diseases associated with DNA transcription, e.g. immunological  
 PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid  
 PT tumours or cancer -  
 XX XX  
 PS Claim 1; SEQ ID No 276; 32pp; English.  
 XX XX  
 CC The invention relates to a nucleic acid, which comprises a segment of the  
 CC chemically pretreated DNA of genes associated with DNA transcription from  
 CC one of 346 sequences, and an oligomer, in particular an oligonucleotide  
 CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical  
 CC to the chemically pretreated DNA of genes associated with DNA  
 CC transcription. The set of oligomer probes are useful for detecting the  
 CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)  
 CC in a chemically pretreated genomic DNA. The nucleic acids are useful for  
 CC diagnosing or treating diseases associated with DNA transcription  
 CC (particularly with the methylation status), e.g. adenosine deaminase  
 CC deficiency, viral infection, retroviral infection, Sezary syndrome,  
 CC haematological disorders, immunological disorders, Werner syndrome,  
 CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,  
 CC neurological disorders, neurodegenerative disorders, Waardenburg  
 CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial  
 CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart  
 CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours  
 CC or cancer. Sequences ABRK28127-ABRK28472 represent DNA transcription  
 CC associated genomic DNA molecules of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format directly from the  
 CC European Patent Office.  
 XX XX  
 SQ Sequence 11047 BP; 3237 A; 248 C; 2358 G; 5204 T; 0 other;





PT Novel nucleic acid useful for diagnosis and therapy of diseases  
PT associated with development genes such as diabetes, comprises a  
PT sequence of a segment of chemically pretreated DNA of genes associated  
PT with development -

PS Claim 1; SEQ ID NO 65; 27bp; English.

XX The invention relates to a nucleic acid (I) comprising a sequence at  
XX least 18 bases in length of a segment of chemically pretreated DNA (II)  
XX of genes associated with development selected from 87 genes listed in  
XX the specification such as ACCPN, ADPN, or AFD1 and comprising one of 350  
XX sequences (ABN7984-ABN8033) or their complements. The invention is  
XX useful for the diagnosis or therapy of diseases associated with  
XX development genes, in particular diseases related to homeobox containing  
XX genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes  
XX associated with congenital heart disease, epilepsy, diseases related to  
XX histone deacetylation, Currarino syndrome, diseases related with the  
XX development of the brain and limb girdle muscular dystrophy and dwarfism.  
XX Oligomers specific to each of the genes are useful for detecting the  
XX methylation state of all CpG dinucleotides within the 350 sequences or  
XX (II) and their complementary sequences, as primer oligonucleotides for  
XX the amplification of the 350 sequences, (II) and/or their complements and  
XX as oligomer probes for detecting the cytosine methylation state and/or  
XX single nucleotide polymorphisms (SNPs).

CC Note: The sequence data for this patent did not form part of the printed  
CC specification but is based on sequence information supplied to Derwent by  
CC the European Patent Office.

XX Sequence 9543 BP; 2652 A; 145 C; 1975 G; 4771 T; 0 other;

Query Match 8.1%; Score 50.2; DB 24; Length 9543;  
Best Local Similarity 46.2%; Pred. No. 0.043;  
Matches 166; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

QY 236 TATTATTATGTTGCTCAAGAGTACTATTATTGAGCCAAATTAATTATTG 235  
DB 2627 TAAGTTTATATTTTTCGTAATTTTAAAGTTTGGTAATATTTTATTTTCG 2686  
QY 296 CTGTACGAGTATTCATTCAGTGTACAGTGTGATGATTAAGCAAGGTT 355  
DB 2687 TGTTTATGATTTATTTATTTAGTTTATTTAAGTGGATTAATTAATTTATTT 2746  
QY 356 TTAATTTAAATGAGTAATTTAATTAATTAATTAATTAATTAATTAATTTGCTTC 415  
DB 2747 TTTTATTTATGTTTGTATTTATTTATTAATTAATTTAATTAATTTATTTATGCT 2806  
QY 416 AACAAACACTTGTTCAGCTTTTACGTTAGTTTCAAGTAATTTGTAAGTTGTC 475  
DB 2807 AGTAATATATTTATTTTAAAGTTGAATATGTTTGTATGTAATGTAATTA 2866  
QY 476 CATATATGATTTAATTAAGAGAGTTTCAACCTTTTACATTAATTAATTAAGAGTCTT 535  
DB 2867 TATATATATATATATATATATATATATTTTGTATTTATTTTATTTATGATGATAT 2926  
QY 536 CAGGTGTGAGATTTATATGCTTAATACAGAAATGACTGTTTATATACGTAAT 594  
DB 2927 TGATTTGTTTATTTTGTGTTAGTTTATTTTGTGTTTGTGTTTGAATTAATGTTCT 2985

RESULT 26  
AA161373/c  
ID AA161373 standard; DNA; 513445 BP.

XX AA161373;

XX 16-OCT-2001 (first entry)

XX Soybean 318013 region A3, SEQ ID NO: 4.

XX Soybean; antihelminthic; gene therapy; soybean cyst nematode; SCN;  
XX SCN resistance; rhg1; Rhg4; SCN resistant allele; plant breeding;  
XX 240017 region G3; 318013 region A3; 515002 region G2; de.

OS Glycine max.

XX WO200151627-A2.

XX 19-JUL-2001.

XX 05-JAN-2001; 2001WO-US00552.

XX 07-JAN-2000; 2000US-0174880.

XX (MONS ) MONSANTO CO.

XX Hauge BM, Wang ML, Parsons JD, Parnell JD;

XX WPL; 2001-425872/45.

XX P-PsDB; AAM42216.

XX New purified nucleic acid for producing a soybean plant having soybean  
XX cyst nematode resistance and for use in plant breeding programs -

XX Claim 30; Page 596-893; 1353bp; English.

CC The invention relates to nucleic acid molecules from regions of the  
CC soybean genome which are associated with soybean cyst nematode (SCN)  
CC resistance. The nucleic acids are used to transform plants, and can  
CC produce soybean plants having an rhg1 or an Rhg4 SCN resistant allele.  
CC The nucleic acids can be used for investigating rhg1 or Rhg4 haplotypes  
CC of soybean plants and for introgressing SCN resistance or partial SCN  
CC resistance into soybean plants. They can also be used in plant breeding  
CC programmes. The invention also relates to proteins encoded by such  
CC nucleic acid molecules, as well as antibodies capable of recognising  
CC these proteins. The present sequence is a nucleic acid molecule  
CC provided in the specification.

XX Sequence 513445 BP; 173367 A; 85402 C; 83912 G; 170492 T; 272 other;

Query Match 8.1%; Score 50.2; DB 22; Length 513445;  
Best Local Similarity 47.4%; Pred. No. 0.099;  
Matches 213; Conservative 0; Mismatches 233; Indels 3; Gaps 2;

QY 147 CGTAATGGACGTCATTTGAAACAAATTTGACATATTTTAA--CAGATCATTTCCG 204  
DB 182573 CGTAGGGGGAATATGATTTTATCAATTTGTATGACTTATTAATTAAGTAATGATA 182514  
QY 205 TTGAGAGTTACAGCAAGTCAGATATGCTTATTTATTTATGTCGTAAGATAGCT 264  
DB 182513 ATTTTACTGGAATTAAGATATATTTAAACATCTTTTACATCAATTTCAAAATCAAAACC 182454  
QY 265 ATTTTATGATGCAATATTTACTTTTATTTGCTGTACGCAATGATTTTCAATGATGTA 324  
DB 182453 TATTATATATGATATGATTAATTTGTAACATATTTTGAACCGGTTGTTTACGAATA 182394  
QY 325 CAAAGTTGTCATTAAGCACTACAAAGTTTAAATTTAATGAAGTAATTTAATATA 384  
DB 182393 AAAAAGACATTTTTCAGATTAATGTTATTAATTAAGATATCAAAATGAAGAAAAA 182334  
QY 385 ATAAAGGGTTGAAGAAAGTTTGTCTTCAACAAACCTTTGAGACTTTTACGTAAT 444  
DB 182333 ATTAAGAAAATTAAGAAATGTAATTAAGAGAAAGAAATGTAAGAAATATTTTAA 182274  
QY 445 TTAGTTCAAGTAATTTGATATGATTTGTTCCATATATGATTAATTAAGAGAGTTTCA 504  
DB 182273 TTTTATATGATTAATGATTAATTAATTAATTAATTAATTTTATTTTATTTTATTT 182215  
QY 505 AACTTTTACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 564  
DB 182214 AATTATGCGTCAATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 182155  
QY 565 GGAATAGTACTGTTTATATATGCTAAT 593  
DB 182154 ATTTTACAAATATTTTCTCTATATTAAT 182126









PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation -

XX Claim 1; SEQ ID NO 583; 32pp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.

XX  
SQ Sequence 18855 BP; 5520 A; 178 C; 3950 G; 9207 T; 0 other;

Query Match 7.8%; Score 48.6; DB 24; Length 18855;

Best Local Similarity 46.9%; Pred. No. 0.12; Mismatches 209; Indels 3; Gaps 1;

Matches 187; Conservative 0; Mismatches 209; Indels 3; Gaps 1;

125 TGGGCTGCTGCTATTTATTCAGCGGTAATGGAACGCTATTTGAAACAATTTCTACTATA 184

10618 TAGGATTAATTTTATTTATTTTATTTAGGGTATTTTATTTTATTTAGTATATTTTA 10677

185 TTTTACGATTCATTTGCTTGGTGAAGTACAGACAGCTGATATTCCTTATTTATTT 244

10678 GTTTTAAAGTTTATTTGTTTGTGTTTAAATATTTGGGTTTATTTTATTTGTT 10737

245 ATGTGCTCAAGAGATGCTATTTATTTATGATGCAATTTATTTATTTGCTGTTACG 304

10738 ATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 10797

305 CATGATTTCTTACGCTGTAACAGTTG--GTGCTATTTGACATCAACAAAGTTTAAAT 361

10798 TTTGTAATGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAAAG 10857

362 TTAATGAGTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 421

10858 TTAATGTTGTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAA 10917

422 CACTGTTGCACTTTTACGATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 481

10918 ATAAAGTTGATTTGTTATTTTATTTAGTTAGTTTAAATTTAAATTTATTTTGG 10977

482 TGATTTATTTAAAGAGAGCTTTCAACTTTTATTTATTTATTTATTTATTTATTTATTT 520

10978 TTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 11016

RESULT 34

ABL92243 ID ABL92243 standard; DNA; 11787 BP.

XX ABL92243;

XX 01-JUL-2002 (first entry)

XX Chemically treated DNA repair gene fragment complementary to #26.

XX DNA repair; cytosine methylation; PMS2L1; PMS2L12; PMS2L2; PMS2L3;

XX PMS2; L4; PMS2L5; PMS2L6; MGMT; MSH2; NUDT1; TDG; INPPL1; RFC4;

XX DDTL1; FANCG; XRCB; ataxia telangiectasia; aging; Bloom's syndrome;

XX Cockayne syndrome; Nijmegen breakage syndrome; Werner syndrome;

XX immunodeficiency; trichothiodystrophy; Fanconi's anaemia; solid tumour;

XX cancer; ds.

XX Unidentified.

XX OS

XX WO200181622-A2.

XX 01-NOV-2001.

XX 06-APR-2001; 2001WO-EP03972.

XX 06-APR-2000; 2000DE-1019058.

XX 07-APR-2000; 2000DE-1019173.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIDEMIOLOGICAL AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-034446/04.

XX Claim 1; SEQ ID NO 52; 25pp + sequence listing; English.

CC The invention relates to nucleic acids containing a sequence of at least

CC 18 nucleotides of chemically treated DNA of genes associated with DNA

CC repair, and their complements. The invention also relates to nucleic

CC acids comprising at least 18 base pairs of the chemically pretreated DNA

CC of genes associated with DNA repair selected from PMS2L1, PMS2L12,

CC PMS2L2, PMS2L3, PMS2, L4, PMS2L5, MGMT, MSH2, NUDT1, TDG, INPPL1,

CC RPA4, DDTL1, FANCG, or XRCB. Nucleic acids of the invention and related

CC oligomers, are useful for diagnosis of diseases associated with gene

CC repair, specifically ataxia telangiectasia, aging, Bloom's syndrome,

CC Cockayne syndrome, Nijmegen breakage syndrome or Werner syndrome,

CC immunodeficiency, trichothiodystrophy, Fanconi's anaemia, solid tumours

CC and/or by detecting single-nucleotide polymorphisms, determination of

CC individual methylation patterns may allow development of individualised

CC therapies. The sequences given in records ABL92152-ABL92335 represent

CC chemically pre-treated DNA fragments from genes associated with DNA

CC repair, and their complements.

CC Note: The sequence data for this patent is not represented in the

CC specification, but is based on sequence information supplied by the

CC European Patent Office.

XX  
SQ Sequence 11787 BP; 3541 A; 134 C; 2217 G; 5895 T; 0 other;

Query Match 7.8%; Score 48.4; DB 24; Length 11787;

Best Local Similarity 46.7%; Pred. No. 0.12; Mismatches 176; Indels 0; Gaps 0;

Matches 154; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

159 TCAATTTGAAACAATTTTCTACTATATTTTAAAGAAATCTGTTGAGGTACAGC 218

4043 TTAATTTGATTTATTTTGTGTTAGTTAGTAAATTTATTTAGTTTATTTATTTATTT 4102

219 ACAGTCAATATGCTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 278

4103 TTTGCTGTTTGAAGCTTTTATTTTAAAGATCGTTTATTTGTTTATTTTATTTTATTT 4162

279 AATATTACTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 338

4163 ATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 4222

339 ATGACTACAAAGCTTTTAAATTTAAATGAGATTAATTTATTTATTTATTTATTTATTT 398

4223 TTTTATTTATTTTGTGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 4282

399 AGAAGCTTTGCTTCAACAAACCTGTTGCACTTTTACGATTTTATTTATTTATTTATTT 458

4283 AATTAATTTATTTTAAAGATGTTATTTGTTATTTTAAAGTTTATTTATTTATTTATTT 4342

459 TGTATAGTATTTGCTCATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 488

4343 TTTAAGTTTATTTAGAGTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 4372

RESULT 35

ABN80017 ID ABN80017 standard; DNA; 19345 BP.

XX AC ABRN0017;  
 XX DT 15-JUL-2002 (first entry)  
 XX DE Human chemically modified disease associated gene SEQ ID NO 34.  
 XX KW Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;  
 KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;  
 KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;  
 KW antidiabetic; cytosatic; anticonvulsant; ds.  
 OS Homo sapiens.  
 OS Synthetic.  
 XX MO200200927-A2.  
 XX PD 03-JAN-2002.  
 XX PF 02-JUL-2001; 2001WO-EP07536.  
 XX PR 30-JUN-2000; 2000DE-1032529.  
 XX PR 01-SEP-2000; 2000DE-1043826.  
 XX PA (EPIG-) EPIGENOMICS AG.  
 XX PI Olek A, Piepenbrock C, Berlin K;  
 XX WP1; 2002-130908/17.  
 XX DR  
 XX PT Novel nucleic acid useful for diagnosis and therapy of diseases  
 PT associated with development genes such as diabetes, comprises a  
 PT sequence of a segment of chemically pretreated DNA of genes associated  
 PT with development  
 XX  
 XX PS Claim 1; SEQ ID NO 34; 27bp; English.  
 XX  
 CC The invention relates to a nucleic acid (I) comprising a sequence at  
 CC least 18 bases in length of a segment of chemically pretreated DNA (II)  
 CC of genes associated with development selected from 87 genes listed in  
 CC the specification such as ACCFN, ADPN, or AFD1 and comprising one of 350  
 CC sequences (ABN79984-ABN80333) or their complements. The invention is  
 CC useful for the diagnosis or therapy of diseases associated with  
 CC development genes, in particular disease related to homeobox containing  
 CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes  
 CC associated with congenital heart disease, epilepsy, diseases related to  
 CC histone deacetylation, Curzaino syndrome, diseases related with the  
 CC development of the brain and limb girdle muscular dystrophy and dwarfism.  
 CC Oligomers specific to each of the genes are useful for detecting the  
 CC methylation state of all CpG dinucleotides within the 350 sequences or  
 CC (II) and their complementary sequences, as primer oligonucleotides for  
 CC the amplification of the 350 sequences, (II) and/or their complements and  
 CC as oligomer probes for detecting the cytosine methylation state and/or  
 CC single nucleotide polymorphisms (SNPs).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification but is based on sequence information supplied to Derwent by  
 CC the European Patent Office.  
 XX  
 XX SQ Sequence 19345 BP; 5413 A; 464 C; 3983 G; 9485 T; 0 other;  
 Query Match 7.8%; Score 48.4; DB 24; Length 19345;  
 Best Local Similarity 45.5%; Pred. No. 0.13;  
 Matches 212; Conservative 0; Mismatches 251; Indels 3; Gaps 1;  
 QY 139 TATTAGGCGTATGGAACGATTTTGAACAATTTCTACTATATTTTACAGATCA 198  
 DB 12502 TTTTATATAGTAGAAGAAAGTTTAAAGTTAATATAGTAATAAGTATGAAATTT 12561  
 QY 139 TTTGCGTTGAGGTACAGACAGTCAGATATGCTTATATTTATTTAGTGTCAAGAG 258  
 DB 12562 TGTGCGATTTAGGTTTATTTAGGTTTAAAGTTGTTGTTTGTGGGAATGTT 12621  
 QY 259 ATAGCTATTTATTTAGATGCATATTAATCTTTTATTTGCTTTACGCGATGATTTCAATTA 318

DB 12622 TTATACGTAATGTTAAAGATGAAATAGATTTTTTTTTTTTCGAAGGAATGAGTT 12681  
 QY 319 CGGTACAGTTGGTGCAATATGACATACAAAGTTTATTAATTAATGAGTAATTT 378  
 DB 12682 ATTATTAATTAAGAGATGATTTGTGTTAAATATATATTTAATTTATTTTATTTAGTA 12741  
 QY 379 AATATTA---TAAAGGTTGAAAGATGTTTCTCTCAACAACACTGTTCGACTT 435  
 DB 12742 TGTATGTTGTTAAGGTTTATTAATTTATTTAATGTAATAATTTAGTTTACG 12801  
 QY 436 TTACGATTTAGTTCAAGTAATTTGTTAAGTATGTTGTTCCATATATGATTAATAAAGA 495  
 DB 12802 AGATTATTTGTTTAAATTTTATTTGTTTGTGTTTATATATATTTTGTGTTT 12861  
 QY 496 GAGTTTCAACCTTTTACATTTATATATATGCAAGTCTCAGGTGGCGAGATTAATG 555  
 DB 12862 AAAAAATTTAGATTAATTAAGTTTGTATGGAATTTAGTTTAATTTATTTATTAAGATA 12921  
 QY 556 CTTAATACAGGAATAGTACTGTTTATATACGCTAATTCCTATGA 601  
 DB 12922 AATTATTTTAAAGGATTAATTTGATATATGATATATTTTATTA 12967  
 RESULT 36  
 ABK40034  
 ID ABK40034 standard; DNA; 12138 BP.  
 XX AC  
 XX AC ABK40034;  
 XX DT 21-MAY-2002 (first entry)  
 XX DE Human chemically pretreated gene sequence #58 strand 2.  
 XX KW Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;  
 KW cytosatic; ALDH6; CYP11A; CYP11B; CYP3A3; DPYD; EPHX2; OCIN; TXNRD1;  
 KW UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.  
 OS Homo sapiens.  
 XX MO200202806-A2.  
 XX PD 10-JAN-2002.  
 XX PF 29-JUN-2001; 2001WO-EP07470.  
 XX PR 30-JUN-2000; 2000DE-1032529.  
 XX PR 01-SEP-2000; 2000DE-1043826.  
 XX PA (EPIG-) EPIGENOMICS AG.  
 XX PI Olek A, Piepenbrock C, Berlin K;  
 XX WP1; 2002-154757/20.  
 XX DR  
 XX PT New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,  
 PT useful for detecting cytosine methylation state of genes associated  
 PT with pharmacogenomics and for therapy of diseases e.g. cancer  
 XX  
 XX PS Claim 1; SEQ ID NO 116; 24bp; English.  
 XX  
 CC The invention relates to a nucleic acid comprising a sequence at  
 CC least 18 bases in length of a segment of the chemically pretreated DNA  
 CC of genes associated with pharmacogenomics according to one of the  
 CC sequences of the genes ALDH6 (NM\_000693), CYP11A (NM\_000781), CYP11B1  
 CC (NM\_000497), CYP3A3 (NM\_000776 and NM\_017460), DPYD (NM\_000110), EPHX2  
 CC (NM\_001979), OCIN (NM\_002538), TXNRD1 (NM\_003330), UGT8 (NM\_003360),  
 CC MRP (NM\_004996, NM\_019900, NM\_019901, NM\_019902, NM\_019862, NM\_019898,  
 CC NM\_019859) and their complementary sequences, or a sequence (SI) chosen  
 CC from 87 sequences and their complements. The chemical pretreatment  
 CC is bisulphite treatment to convert cytosines (but not methyl-cytosines)  
 CC into uracils. Also included are an oligomer (II) in particular an  
 CC oligonucleotide or a peptide nucleic acid (PNA)-oligomer, comprising in





OS Homo sapiens.  
 XX WO200200928-A2.  
 XX  
 XX 03-JAN-2002.  
 XX  
 XX 02-JUL-2001; 2001WO-EP07537.  
 XX  
 XX 30-JUN-2000; 2000DE-1032529.  
 XX PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 XX (EPIC-) EPIGENOMICS AG.  
 XX  
 XX Olek A, Piepenbrock C, Berlin K;  
 XX WPI; 2002-130909/17.  
 XX  
 XX Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation -  
 XX  
 XX Claim 1; SEQ ID NO 1893; 32bp + Sequence Listing; German.  
 XX  
 XX The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 CC  
 XX Sequence 12590 BP; 2516 A; 310 C; 3233 G; 6531 T; 0 other;  
 SQ  
 Query Match 7.7%; Score 48.2; DB 24; Length 12590;  
 Best Local Similarity 45.0%; Pred. No. 0.13;  
 Matches 220; Conservative 0; Mismatches 268; Indels 1; Gaps 1;  
 Oy 80 AATCAGAGAGGTTCTAAGATGACTGACGACGAGGATGCTGGGCTGCTATT 139  
 Db 752 ATTATAGTATTAATTAATTAAGATGATTTGTTATGTTATGATTTAGG 871  
 Oy 140 ATTCAGCGGTATGCGACGCTCATTTTGAACAATTTCTACTATATTTTACAGATCAT 199  
 Db 812 AATTAGTGTATTTGGGTAATAAATTAGAGAAATTTGTTTGTATTTTATGATTTAGG 871  
 Oy 200 TTCGGTTGAGTTACACACAGTCAGTATGCTTTAT-TTATTTATGCTGCTCAAGG 258  
 Db 872 GATTGTTTGTATTTTATTTGTTGTTATTAATTTTATTTTATGAGCGTTTGG 931  
 Oy 259 ATTAGCTATTTATGATGCAATATTTACTTTTATGCTGTTAGCGCAGTATTCATTA 318  
 Db 932 GAATTTTGTGTTTATGTTATTAATTTTTCGTGTTGTTAAATTTTGTGTTG 991  
 Oy 319 CGTGCACAGTTGTCATTAATGACTACAAAGTTTAAATTTAAATGAGATTAATTT 378  
 Db 992 TTTTATTTTGAAGAGATTAAGTTAGTTATGTTTATTTTAAATTTTATTAATA 1051  
 Oy 379 AATATATATAAAGGTTGAAGAAGATGTTGCTTCTCAACAAACCTGTTGACTTTTA 438  
 Db 1052 TTTGTAATAAATTTTATTTTAAATAGATTAATTTATGTTTGAAGATTTTGT 1111  
 Oy 439 CGTAGTTAGTTCAAGTATTTAGTATTTGTTCCATTAATTAATTAATAAAGAGAG 498  
 Db 1112 GGGGGTTTGTATTAATTAAGTTTGAATGCTTATATTTTGTGTTTATTAATGAATTA 1171  
 Oy 499 TTTTCAACCTTTTACATTAATTAATTAATGCAAGTCCCTGAGGTGCGAGATTAATGCTT 558  
 Db 1172 TTTATTAAGTTTGTGTTTATTTTATTTTGAATATTTGTAATGTTTAAATGTTGAA 1231  
 Oy 559 AATACAGA 567  
 Db 1232 TAGTAGGA 1240

RESULT 39  
 ABL32977  
 ID ABL32977 standard; DNA; 18512 BP.  
 XX  
 XX ABL32977;  
 XX  
 XX 26-MAR-2002 (first entry)  
 XX  
 XX Human immune system associated gene SEQ ID NO: 950.  
 XX  
 XX Human; immune system disease; cytosine methylation; antiasthmatic;  
 XX antiarteriosclerotic; antiamebic; cytostatic; nocotropic;  
 XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 XX antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 XX antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 XX gene; ds.  
 XX  
 XX OS Homo sapiens.  
 XX  
 XX WO200200928-A2.  
 XX  
 XX 03-JAN-2002.  
 XX  
 XX 02-JUL-2001; 2001WO-EP07537.  
 XX  
 XX 30-JUN-2000; 2000DE-1032529.  
 XX PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 XX (EPIC-) EPIGENOMICS AG.  
 XX  
 XX Olek A, Piepenbrock C, Berlin K;  
 XX WPI; 2002-130909/17.  
 XX  
 XX Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation -  
 XX  
 XX Claim 1; SEQ ID NO 950; 32bp + Sequence Listing; German.  
 XX  
 XX The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 CC  
 XX Sequence 18512 BP; 5520 A; 92 C; 3489 G; 9411 T; 0 other;  
 SQ  
 Query Match 7.7%; Score 48.2; DB 24; Length 18512;  
 Best Local Similarity 49.5%; Pred. No. 0.14;  
 Matches 152; Conservative 0; Mismatches 153; Indels 2; Gaps 1;  
 Oy 168 AACAAATTTCTACTATATTTTACAGATCAATTTGCGTTGAGTTACAGACACTCAGT 227  
 Db 4706 AATATTAAGTAATTTAGTTGAATGATTAATTTGTTTGAAGTATTAATAATTTTGT 4765  
 Oy 228 ATATGCTTATTTATTTATGTTGCTCAAGATAGCTATTTTATTTAGCCAAATTTACT 287  
 Db 4766 TTTTATATTTGTTTATTTATTTTATTTTATGTTTATTTATTTTATTTATTTATTT 4825  
 Oy 288 TTTTATGCTGTTACGCGATGATTTCTATGCTGTAACAAGTTGCTGATTAAGACTAC 347  
 Db 4826 GTGATTTGTTATTAATTAAGATTAAGATGCTTTTATTAATGTTTATTTATTTATG-1GG 4883  
 Oy 348 AAAGTTTTTAATTAATTAAGATTAATTTAATTAATTAATAAAGGTTGAAGAAGATGTT 407

Db 4884 GAAAGGAAAGAGATATATAATTAATAATATATATGTTTATTTATTTATTTA 4943  
 QY 408 TGGTTCACACAAACACTTGTGACCTTTACGTAAGTTAGTCAAGTATTTGTAATAG 467  
 Db 4944 TGTAGTATTTTATTTGAGACTTTTATTTTATTTATTAATTAATTAATGTTTAATG 5003  
 QY 468 TATTTGTT 474  
 Db 5004 TTTTTTTT 5010

## RESULT 40

ABN85766/c  
 ID ABN85766 standard; cDNA; 82952 BP.

AC ABN85766;

DT 21-OCT-2002 (first entry)

DE Arabidopsis yellow stripe1-like 3 encoding cDNA SEQ ID NO 7.

KW Maize; transgenic; plant; yellow stripe1-like; ysl; Arabidopsis;

KW Iron uptake; bioremediation; yellow stripe 1; ysl; gene; ss.

OS Arabidopsis sp.

PN MO200240688-A2.

PD 23-MAY-2002.

PF 16-NOV-2001; 2001WO-US43101.

PR 16-NOV-2000; 2000US-249222P.

PA (UYTA ) UNITV YALE.

PI Walker EL, Dellaporta S;

DR WPI; 2002-480144/52.

DR P-PSDB; ABB83919.

PT New yellow stripe1 and yellow stripe1-like genes, useful for altering  
 the distribution of iron within the plant body so that edible parts of  
 crop plants have more iron, or for producing plants useful in enhancing  
 iron uptake from soil

PS Claim 1; Page 65-111; 187bp; English.

CC The invention relates to an isolated nucleic acid molecule (I), maize  
 CC yellow stripe 1 (ysl) or yellow stripe-like (ysl) from Arabidopsis  
 CC (ABN85763-ABN85771). (II) is useful for generating transgenic plants which  
 CC can be used for enhancing iron uptake from soil and for bioremediation of  
 CC metal or heavy metal contaminated soil. (I) may also be used to alter the  
 CC distribution of iron within the plant body so that edible parts of crop  
 CC plants have more iron. Transgenic plants may also be used in conventional  
 CC plant breeding schemes to produce progeny which also contain the gene of  
 CC interest. The present sequence is that of the Arabidopsis ysl encoding  
 CC cDNA of the invention.

XX SQ Sequence 82952 BP; 26971 A; 14261 C; 14365 G; 27355 T; 0 other;

XX Query Match 7.7%; Score 48.2; DB 24; Length 82952;

XX Best Local Similarity 48.4%; Pred. No. 0.2; Mismatches 143; Indels 0; Gaps 0;

QY 281 TATTAATTTTATTTGCTGTACGCAATGATTCATTACGTCGTACAGTTGGTCATTAT 340

Db 21657 TACTAGCTGCTGTGATCATATTTGGGTTTCTATTTGTTATTAAGCTACTGAACATT 21598

QY 341 GGACTCAAAAGGTTTAAATTAATGAATTAATTAATTAATTAATTAATTAATTAATTAAT 400

Db 21597 CGACCAACATCTCATTTAAACCTTTCAAAATAGATACAAACCTAGAAAGAAAG 21538

QY 401 GAATGTTGCTTCTCAACAAACACTTGTGCACTTTTACGTAAGTTAGTCAAGTATTTG 460  
 Db 21537 CATCATATCTCTATCATCAAAACCTTTATCATTAATAATTTGGTAATTAAGAAATCT 21478  
 QY 461 TTATAGGTAATGTTCCATATATATGATTAATAAGAGAGTTTCAAACTTTTACATTAT 520  
 Db 21477 TTAACTCTGCTTATCAGAAATTTTAAATGAAGATCATCATATATGCTTTTAT 21418  
 QY 521 ATTATGCAAGTCTTCAGAGTGTGCAAGATTAATATGCT 557  
 Db 21417 TAGATATATGCTTTTAAATATGCTTTCTCAATCT 21381

Search completed: October 4, 2003, 04:04:50  
 Job time : 223 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2003, 03:48:37 ; Search time 56 Seconds  
(without alignments)  
4902.506 Million cell updates/sec

Title: US-10-010-160-1  
Perfect score: 622  
Sequence: 1 atgctgatgagccagctaa.....aattatcgagtcgcagatc 622

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	46.6	7.5	8920	2 US-08-446-855A-1	Sequence 1, Appli
C 2	46.6	7.5	8920	3 US-09-150-741-1	Sequence 1, Appli
C 3	44.4	7.1	5181	1 US-08-257-073-10	Sequence 10, Appli
4	44.4	7.1	5852	1 US-07-867-106-2	Sequence 2, Appli
5	43.8	7.0	70000	4 US-09-851-896-3	Sequence 3, Appli
6	43.6	7.0	19124	2 US-08-487-826B-13	Sequence 13, Appli
7	42.8	6.6	66804	4 US-09-740-041-3	Sequence 3, Appli
C 8	41	6.9	3255	4 US-09-601-198-108	Sequence 108, App
C 9	40	6.4	19124	2 US-08-487-826B-13	Sequence 13, Appli
10	39.8	6.4	1400	2 US-08-305-764C-57	Sequence 57, Appli
11	39.8	6.4	1400	2 US-08-305-764C-55	Sequence 55, Appli
C 12	39.8	6.4	2095	2 US-08-305-764C-55	Sequence 55, Appli
C 13	39.8	6.4	9636	1 US-08-323-170B-1	Sequence 1, Appli
C 14	39.8	6.4	9636	1 US-08-459-415-1	Sequence 1, Appli
15	39	6.3	4526	1 US-07-855-412B-4	Sequence 4, Appli
16	39	6.3	4526	2 US-08-308-887A-4	Sequence 4, Appli
17	39	6.3	4526	3 US-08-881-094-4	Sequence 4, Appli
18	38.8	6.2	2861	1 US-08-299-953-1	Sequence 1, Appli
19	38.8	6.2	2861	1 US-08-459-415-1	Sequence 1, Appli
20	38.8	6.2	2861	4 US-09-066-687-1	Sequence 1, Appli
21	38.8	6.2	2861	5 PCT-US95-11231-1	Sequence 1, Appli
22	38.8	6.2	3881	1 US-08-299-953-2	Sequence 2, Appli
23	38.8	6.2	3881	1 US-08-459-415-2	Sequence 2, Appli
24	38.8	6.2	3881	4 US-09-066-687-2	Sequence 2, Appli
25	38.8	6.2	3881	5 PCT-US95-11231-2	Sequence 2, Appli
26	38.6	6.2	1664976	4 US-08-916-421B-1	Sequence 1, Appli
27	37.8	6.1	701	3 US-08-998-416-701	Sequence 701, App

C 28	37.8	6.1	1401	4 US-09-134-001C-595	Sequence 595, App
C 29	37.8	6.1	2142	4 US-09-107-532A-905	Sequence 905, App
C 30	37.8	6.1	2434	4 US-09-489-847-67	Sequence 67, Appli
31	37.6	6.0	1507	1 US-07-641-143B-1	Sequence 1, Appli
32	37.6	6.0	1507	1 US-08-124-290-1	Sequence 1, Appli
33	37.6	6.0	1507	3 US-08-686-372A-1	Sequence 1, Appli
34	37.4	6.0	1296	4 US-09-107-532A-3402	Sequence 3402, Ap
35	37.4	6.0	1875	1 US-08-286-325A-1	Sequence 1, Appli
36	37.4	6.0	3234	4 US-08-286-325A-7	Sequence 2508, Ap
37	37.2	6.0	576	4 US-09-134-001C-2508	Sequence 3, Appli
C 38	37.2	6.0	6124	4 US-08-213-419B-3	Sequence 7, Appli
39	37	5.9	1918	4 US-09-599-360B-7	Sequence 60, Appli
40	37	5.9	1918	4 US-09-599-360B-60	Sequence 60, Appli
C 41	37	5.9	8501	3 US-09-298-367B-6	Sequence 595, App
42	36.8	5.9	658	3 US-08-998-416-595	Sequence 1, Appli
C 43	36.8	5.9	6152	3 US-08-973-462-1	Sequence 1, Appli
44	36.6	5.9	767	3 US-08-998-416-472	Sequence 472, App
45	36.4	5.9	3618	1 US-07-872-678A-36	Sequence 36, Appli

ALIGNMENTS

RESULT 1  
US-08-446-855A-1/c  
Sequence 1, Application US/08446855A  
Patent No. 5849573  
GENERAL INFORMATION:  
APPLICANT: Stewart, Thomas S  
APPLICANT: Flores, Maria V  
APPLICANT: O'Sullivan, William J  
TITLE OF INVENTION: Nucleotide sequence encoding carboxyl  
TITLE OF INVENTION: phosphate synthetase II  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon & Vanderhye PC  
STREET: 1100 No. 5849573th Glebe Road, 8th Floor  
CITY: Arlington  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/446, 855A  
FILING DATE: 06-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mitchard, Leonard C  
REGISTRATION NUMBER: 29,009  
REFERENCE/DOCKET NUMBER: 47-80  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-816-4000  
TELEFAX: 703-816-4100  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8920 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: genomic  
US-08-446-855A-1  
Query Match  
Best Local Similarity 7.5%; Score 46.6; DB 2; Length 8920;  
Matches 211; Conservative 0; Mismatches 214; Indels 6; Gaps 3;  
OY 173 TTTTCTACTATATTTTACAGATCATTCGTTGAGCTTACAGACAGTCAGATATG 232  
DB 8835 TTTCGTGAAATTTTAAATAAATCATTAACATTATTAATTTATATTAATTAAGAAT 8776

Oy	223	CTTATATTATTTATGCTGCTCAAGAGATAGCTATTTATTTGATGCCAATATCTTTTA	292
Db	8775	TTATRCATTTTAAATGTTATATATTTTATTTTATTTTATTAATTAATTTATTTAAAT	8716
Oy	293	TTGCTGTTACGGCAGATTCATTACGTGACAAAGTGGTGCAATATGACATACAA	-G 351
Db	8715	TATAAAATTAATTTGPAATATGAAAAACCATTTTGGTTATACATATGATGAATPAATA	8656
Oy	352	GTTTAAATTTTAAATGAGATTAATTTATATATATTAATTAAGGTTGAAGAAATGTTTCT	411
Db	8655	AAATTTGATATATATACAAAAATTTATTAATAAAAAATTCATATPAATTTATGATCATTA	8596
Oy	412	TCTCAACAAACACTGTTGCACTTTTA---CGTACTTATAGTTCAGATATTTGTTAGGT	468
Db	8595	TTTATATAAAAACATTATTTAATTTTAAACGTAATATATATTAATTAATTAATTAATATAT	8536
Oy	469	ATTGTTCAATATGATTTATTAATAAGAGAGTTTCAAACTTTTACATATATATATGCA	528
Db	8535	ATCA	8476
Oy	529	AGTCTTCAAGGTGCGCAG--ATTATATGCTTATATACGAATAGTACTGTTTATATA	586
Db	8475	TGTTGTTATGCAATGAAAGTTATATATAATACATAAATAAGTATGTTAAATCGAATTATAT	8416
Oy	587	CGCTAATTCCT 597	
Db	8415	GAATATATTTCT 8405	

## RESULT 2

US-09-150-741-1/c  
; Sequence 1, Application US/09150741

```

GENERAL INFORMATION:
APPLICANT: Stewart et al.
TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
Patent No. 6183996
TITLE OF INVENTION: Synthetase II
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/09/150,741
CURRENT FILING DATE: 1998-09-10
EARLIER APPLICATION NUMBER: PL6380
EARLIER FILING DATE: 1993-12-16
EARLIER APPLICATION NUMBER: AU93/00617
EARLIER FILING DATE: 1993-12-02
EARLIER APPLICATION NUMBER: 08/446,855
EARLIER FILING DATE: 1995-07-06
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 8920
TYPE: DNA
ORGANISM: Plasmodium falciparum
US-09-150-741-1

```

Query Match : 7.5%; Score 46.6; DB 3; Length 8920;

Best Local Similarity 49.0%; Pred. No. 0.011;  
Matches 211; Conservative 0; Mismatches 214; Indels 6; Gaps 3,

Qy	173	TTTCTACATAATTTTTCAGAAATCATTTCCGTTTGAGGTTCACACACAGTCAGTAATG	232
Db	8835	TTTCGTGAAATTTTTTAAAAAATCATTAACATTATTAATTTATATATATTTAAAAAGAT	8778
Qy	233	CTTATATTATTTATGTTGCTCAAGATAGCTATTTTATGTAGTCCAAATATTACTTTTAA	292
Db	8775	TTATTCATTTTAAAGTTTATATATTTTATTTTTTTTATTAATTAATTAATTTTAAAT	8712
Qy	293	TTGCTGTACGGCATGGAATTCATTACGTACAAAGTGGTCATTATGAGCTACAAA-G	351
Db	8715	TATAAATTAATTTGTAAATATGAAAAAACCATTTTGGTTATACATATAGTGATTAATAA	865
Qy	352	GTTTTAAATTTAAATGAGCTAAATTTATATTAATTAATAAAGGTTGAAGCAATGTTTCT	411

Position	Sequence	Position	Sequence
Db	8655 AAATTGTGATATATAACAAAATTTATTAATAAATAATCAATTAATTAATGATCAT	8596	
Oy	412 TCTCAACAAACACGTTGTTCGACTTTA---CGTAGTTAGTCAAGTAATGTATAGT	468	
Db	8595 TTATATATAAACATTATTAATATATTTAAACGTAAATTAATTAATAATATATAT	8536	
Oy	469 ATTGTTCATATATGATTAATATAAAGAGAGTTTCAACCTTTTACCATTAATATATGCA	528	
Db	8535 AT	8476	
Oy	529 AGTCCCTCAGGTGGGAG--ATATATGCTTAATACAGAAATGTACTGTTTATATA	586	
Db	8475 TGTGTTATGGAATGAAGAAGTAATTAATATACATTAATAGTATGTAAATGCAATTTAT	8416	
Oy	587 CGCTAATTCCT	597	
Db	8415 GAATATATTTCT	8405	

### RESULT 3

US-08-257-073-10/c  
; Sequence 10, Application US/08257073

GENERAL INFORMATION:

APPLICANT: de Taisne, Charles

TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE

CORRESPONDENCE ADDRESS:

STREET: 530 Fifth Avenue, 25th Floor

STATE: New York  
COUNTRY: UNITED STATES OF AMERICA

ZIP: 10036

MEDIUM TYPE: Floppy, disk  
COMPILED: IBM PC: csmc34-ibj

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatonTr Release #1.0

! CURRENT APPLICATION DATA: .  
APPLICATION NUMBER: US/09/257 073

FILING DATE: 09-JUN-1994  
CLASSIFICATION: 42A

PRIOR APPLICATION DATA: IIS 08/075 783  
APPLICATION NUMBER:

FILING DATE: 11-JUN-1993  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/852,305  
FILING DATE: 18-MAR-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IIS 07/673 193

FILING DATE: 20-MAR-1991

NAME: Frommer, William S.  
REGISTRATION NUMBER: 35 506

REFERENCE/DOCKET NUMBER: 454310-2570

TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712

TELEX: 425066 CURTMS  
INFORMATION FOR SEQ ID NO. 10-

SEQUENCE CHARACTERISTICS:

TYPE: nucleic acid

! TOPOLOGY:  
ms-08-257-073-10

Query Match	7.1%;	Score 44;	DB 1;	Length 5181;
Best Local Similarity	53.5%;	Pred. No. 0.042;		

	Matches	92:	Conservative	0:	Mismatches	80:	Indels	0:	Gaps	0:
QY	404	TGTTTGCCTTCACAAACAACTGTTGCACCTTTAGCTAGTTAGTTCAGTAATGTTTA	463							
Db	2733	TTTTTGCCTCCMCATAATATTTTTCATGTGATTTCTCTTTAGTTATGTAATAATA	2674							
QY	464	TAGGATTTGTTCCATATATGATTTAATAAAGAGAGCTTTCAACTTTTACCATTATATT	523							
Db	2673	AATCATTTCTTTTGATATTAATTCAAAAAGAGATGTGTAATCATTTGTCATACATATC	2614							
QY	524	ATGCAGCTCTTCAGGTGTGCAGATTTATATGCTTAATACAGGAATGACT	575							
Db	2613	ATAATATATCAATACAGCAGTATGTTATTTTGAATATTAATATTAATCT	2562							

```

1  RESULT 4
2  US-07-867-106-2
3  / Sequence 2, Application US/07867106
4  / Patent No. 5389526
5  / GENERAL INFORMATION:
6  / APPLICANT: Slade, Martin B
7  / APPLICANT: Chang, Andy C M
8  / APPLICANT: Williams, Keith L
9  / TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
10 / TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
11 / NUMBER OF SEQUENCES: 19
12 / CORRESPONDENCE ADDRESS:
13 / ADDRESS: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
14 / STREET: One Liberty Place 46th Floor
15 / CITY: Philadelphia
16 / STATE: PA
17 / COUNTRY: USA
18 / ZIP: 19103
19 / COMPUTER READABLE FORM:
20 / MEDIUM TYPE: Floppy disk
21 / COMPUTER: IBM PC compatible
22 / OPERATING SYSTEM: PC-DOS/MS-DOS
23 / SOFTWARE: Patentin Release #1.0, Version #1.25
24 / CURRENT APPLICATION DATA:
25 / APPLICATION NUMBER: US/07/867,106
26 / FILING DATE: 19920625
27 / PRIOR APPLICATION DATA:
28 / APPLICATION NUMBER: AU PJ 7187
29 / APPLICATION NUMBER: PCT/AU90/00530
30 / FILING DATE: 02-NOV-1989
31 / ATTORNEY/AGENT INFORMATION:
32 / NAME: Feeney, Joanne Longo
33 / REGISTRATION NUMBER: 35,134
34 / REFERENCE/DOCKET NUMBER: RICE-0002
35 / TELECOMMUNICATION INFORMATION:
36 / TELEPHONE: 215-568-3100
37 / TELEFAX: 215-568-3439
38 / INFORMATION FOR SEQ ID NO: 2:
39 / SEQUENCE CHARACTERISTICS:
40 / LENGTH: 5852 base pairs
41 / TYPE: NUCLEIC ACID
42 / STRANDEDNESS: single
43 / TOPOLOGY: linear
44 / MOLECULE TYPE: DNA (genomic)
45 / ANTI-SENSE: NO
46 / FEATURE:
47 / NAME/KEY: CDS
48 / LOCATION: 2378..5038
49 / FEATURE:
50 / NAME/KEY: CDS
51 / LOCATION: 2378..5038
52 / US-07-867-106-2
53
54 Query Match 7.1%; Score 44; DB 1; Length 5852;
55 Best Local Similarity 47.2%; Pred. No. 0.043;
56 Matches 134; Conservative 0; Mismatches 150; Indels 0; Gaps 0
57
58 311 TTGCTATTCAGTGTACAAAGTGGTGCATTATGCACTACAAAGGTTTAAATTTAAATGGA 370

```

Db	2088	TTTTTAATTAATTTAAATTTTATTTCTGATCTATATATACCTTATTTATTTAAATTTGGAT	2147
Qy	371	GTAAATTTAATATATATATAAAGGGTTGAAGGAATGTTGGCTTCTCAACAAACACTGTTC	430
Db	2148	AATATATCAAAATATTTATATCAGTTTGGCATGACAAATTTAATATATATTTATTTTGGATT	2207
Qy	431	GACTTTTACGTAGTTTAGTTCAGTAAATGTTATAGTATTTGTTCCATATATAGTATTA	480
Db	2208	AATTTTTTTTTTTTTTTTTTTTTTAAATTTCTTTTTTTTTTTTTTTTATTTTAAATTTTAA	2267
Qy	491	AAGAGAGTTTTTCAACTTTTTACCATTAATATATATGCAAGTCCTCAGTGGCAGATT	550
Db	2268	ATTTTTATTTTTTCCCACTTTCATTTTATTTTATTTTATTTATTTATTTGTAATATCATTTTAT	2327
Qy	551	ATAAGCTTAATACGAATAAGTACTGTTTATATATACGTAAAT	594
Db	2328	TTATTTTAAATTAATATGTTTGGTTTAATTTTATTTCAAAAGATT	2371

```

RESULT 5
US-09-851-896-3
; Sequence 3, Application US/09851896
; Patent No. 6410325
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Susan M. Freier
; APPLICANT: Andrew T. Malt
; TITLE OF INVENTION: ANTISENSE MODULATION OF PHOSPHOLIPASE A2, GROUP VI (CA2+-INDEPENDENT)
; TITLE OF INVENTION: EXPRESSION
; FILE REFERENCE: RTS-0220
; CURRENT APPLICATION NUMBER: US/09/851,896
; CURRENT FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 3
; LENGTH: 70000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; US-09-851-896-3

Query Match          7.0%; Score 43.8; DB 4; Length 70000;
Best Local Similarity 46.2%; Pred. No. 0.098;
Matches 199; Conservative 0; Mismatches 217; Indels 15; Gaps 1;

Cy 183 TATTTTACGAATTCATTGGTGGTTGAGTGTACAGCAGACGTCGATATGCTTATTTAT 242
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 23179 TAGTTTGTGTTTAATGTTGTGTAATAATTAGCAGTGACGCCACTGGGTCCAGGCTTT 23238

Cy 243 TTATGTTGCTCAAGAGATAGCTATTTTATTTGATGCCAATTTACTTTTATGCTGTAC 302
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 23239 CTTTGCGGGAGACTTTTCATTACAGCTTGATCTGATTACTTGTTATGTCATTATAG 23298

Cy 303 GGCATGATTTTCATTACGTCGTACAGTGGTGCAATTATGACATACAAAGSTTTTAAAT 362
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 23299 GTTTTGATTTCTTCACAGGTTCCATCTTGAGTGTGTATGCTTAGCAATTTATCCAT 23358

Cy 363 TAAATGAGTAATTTATATATATATAAAGGTTGAAAGAAAGTGTGCTCTCAACAAAC 422
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 23359 TTCTTTAGAGGTTTTCATTATATGGCA-----TATATGTTTCACAGTA 23403

Cy 423 ACTTGTCGACTTTTACGTAGTTAGTTCAGATTAATGTTATAGTAATGTTCCATATAT 482
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 23404 ACCCTAATATATCCTTTGAGATTTCTGTGATATTTGGTTGTAATATCTTTTGTCAATTC 23463

Cy 483 GATTATTAAGAAGAGAGTTTCAACCTTTTACCATATATATATGCAAGTCTTCAGGTGT 542
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 23464 GATTTTGTGTTGGGGTTCCTCTCTTTTGTGTAGTCAAGTAAGATTTATGATTT 23523

Cy 543 GGCAGATTATATGCTTATATACAGGAATAGACTTGTTTATATATACGCTAATTCCTATGAC 602
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 23524 TGTTTATCTTTTCAAAAACCTTATATTTGTGTATCTTTGTATGTTGTTTGTTCAAATTC 23583

```





```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 108
; LENGTH: 3255
; TYPE: DNA
; ORGANISM: Ureaplasma urealyticum
US-09-601-198-108

```

Query Match	6.6%;	Score 41;	DB 4;	Length 3255;
Best Local Similarity	47.5%;	Pred. No. 0.21;		
Matches 122;	Conservative 0;	Mismatches 135;	Indels 0;	Gaps 0

QY	321	TGTCACAGTTGGTCATATATGACATCAACAAGCTTTTAAATTTAATGAGATAATTTVA	380
Db	1645	TTTCATTAACCTACATTAAAAAGATTAATCCAGATTGATCTTAAAAAAGATTAAGATA	1588
QY	381	TATATAAAAAGGTTGAAAGGATGTTGGCTTCCACAACAACCTGTTGACTTTACG	440
Db	1585	TTATTTTCAATTTACAAAAAATTTAAATAGATTAAAAATTTAAATCTGTGTAATGAATA	1528
QY	441	TAGTTTAGTCAAGTAATGTATATAGTATGTTTCCATATATGATTAATAAAGAGAGTT	500
Db	1525	AAATTTAGCTTTTATATATTACTCTTTGGTTTATGATCCAAATCTCATCAAGAAACTA	1466
QY	501	TTCAACCTTTTACCAATTATATATATGCAAGTCCTTCAGGTGGCAGATTATATGCTTA	560
Db	1465	GTGAAAAAACCTTATGTTAAGGTCCAATTTCAATAGACGTCCCTTAAAAACTTAA	1406
QY	561	TACAGAAATAGTACTTG	577
Db	1405	AAATTAATAAAGTTATG	1389

RESULT 9  
US-08-487-826B-13/c

; Sequence 13, Application US/08487826E  
; Patent No. 6003037

GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chlcnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:

ADDRESS: Knodde Martens Olson & Bear  
 STREET: 620 Newport Center Drive 16th Floor  
 CITY: Newport Beach  
 STATE: California  
 COUNTRY: US  
 ZIP: 92660  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.2  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/487, 822

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-487-826B-13

```

Query Match	6.4%	Score 40;	DB 2;	Length 19124;
Best Local Similarity	45.1%;	Pred. No. 0.63;		
Matches 148;	Conservative 0;	Mismatches 180;	Indels 0;	Gaps 0;

Oy	161	ATTTTGAACAATTTTTCACIATATTTTACGAATCATTTGGGTGAGTTACGAC	220
Db	15844	ATTTTTTATTTATTTATTTTATTTTAAATTTTATTTATTTATGATATAT	157685
Oy	221	AGTCAGATATGCTTATTTATTTATGTTGTCACAGATPACGATTTTATTTGATGCCA	280
Db	15784	TTTTTTTTTAACTTTTTTTTAAATTTTTTTTTTAAATTTTATGATATATTTTTTAAATAT	157295
Oy	281	TATTACTTTTATTTGGCGTTACGGCAGATTTCTACGTACAAAGTTGGCATAT	340
Db	15724	ATTTTTCCTTTTTTTTTTTGTTTATGATATATTTTTTTTTTTTAAATGTTTTTTTT	156655
Oy	341	GGACTACAAAGTTTTTAAATTTAAATGAGTAAATTTAATATTAATAAAAGGTTGAAG	400
Db	15664	TTTCTCTTTTGTTTTTATTTTTTTTATATCATTTTTTTTTTAAATATAAAATTTTTTTTT	156050
Oy	401	GAAGTTGGCTTCACAAACACTGGTCGACTTTACGTAGTTAGTTCAGATATG	460
Db	15604	AATTTTTTTTGAATATCTTTTGATTTTATTTATCTATCAAAATTTATTTTATATTA	155454
Oy	461	TTATAGTATGTTCCATATATGATAT	488
Db	15544	TTTTTATTTATTTTTTAAAAATTTTTCT	15517

RESULT TO  
US-08-305

; Sequence '57, Application US/08305764C  
; Patent No. 5856090

GENERAL INFORMATION: David M.  
APPLICANT: Epstein, David M.  
TITLE OF INVENTION: DNA METHYLASE LINKING REACTION  
NUMBER OF SEQUENCES: 71  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE  
STREET: 10550 No. 5656090th Torrey Pines Road  
CITY: La Jolla  
STATE: California  
COUNTRY: USA  
ZIP: 92037  
COMMUNICATED BY: COMM.

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0. Va  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/305,764C  
FILING DATE: 09-SEP-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitting, Thomas  
REGISTRATION NUMBER: 34,163  
REFERENCE/DOCKET NUMBER: TSRI 440.0.

MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1257  
US-08-305-764C-57

Query Match 6.4%; Score 39.8; DB 2; Length 1400;  
Best Local Similarity 48.5%; Pred. No. 0.34;  
Matches 110; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

OY 308 GGATTCATTACGTGACAAAGTTGGGACATTATGAGCTACAAAGTTTAAATTAAT 367  
DB 887 GCATTCAGACATTACAAAAGATTATCTTTAAAGATGATGTAACCTTCT 946  
OY 368 GGAGTAATTTAATTAATAAAGGTTGAAAGATGTTGCTTCAACAAACCTTG 427  
DB 947 TAATTGACAAAATACGACTGGGACGTTAAACATTGTTCTACATCAAAAATTC 1006  
OY 428 TTCGACTTTACGTAGTTAGTTCAAGTAATGTTATAGGATGTTCCATATATGATTA 487  
DB 1007 AACGATTAACAGGTACTTTGTTAAGATGAGAAACAGATACCGTCTTTAACAACGA 1066  
OY 488 TAAAGGAGAGTTTCAACCTTTTACCATTAATATGCAAGTCT 534  
DB 1067 ATGAATGCAAGCTATTATGGGTTTCCAAAAGATTTGTATTCT 1113

## RESULT 11

US-08-305-764C-59  
Sequence 59, Application US/08305764C  
Patent No. 5856090

GENERAL INFORMATION:

APPLICANT: Epstein, David M.

TITLE OF INVENTION: DNA METHYLASE LINKING REACTION

NUMBER OF SEQUENCES: 71

CORRESPONDENCE ADDRESS:

ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE

STREET: 10550 No. 5856090th Torrey Pines Road

CITY: La Jolla

STATE: California

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/305,764C

FILING DATE: 09-SEP-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Fitting, Thomas

REGISTRATION NUMBER: 34,163

REFERENCE/DOCKET NUMBER: TSRI 440.0

TELEPHONE: (619) 784-2937

TELEFAX: (619) 784-9399

INFORMATION FOR SEQ ID NO: 59:

SEQUENCE CHARACTERISTICS:

LENGTH: 1400 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ANTI-SENSE: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1257

US-08-305-764C-59

Query Match 6.4%; Score 39.8; DB 2; Length 1400;  
Best Local Similarity 48.5%; Pred. No. 0.34;  
Matches 110; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

OY 308 GGATTCATTACGTGACAAAGTTGGGACATTATGAGCTACAAAGTTTAAATTAAT 367  
DB 887 GCATTCAGACATTACAAAAGATTATCTTTAAAGATGATGTAACCTTCT 946  
OY 368 GGAGTAATTTAATTAATAAAGGTTGAAAGATGTTGCTTCAACAAACCTTG 427  
DB 947 TAATTGACAAAATACGACTGGGACGTTAAACATTGTTCTACCTATCAAAAATTC 1006  
OY 428 TTCGACTTTACGTAGTTAGTTCAAGTAATGTTATAGGATGTTCCATATATGATTA 487  
DB 1007 AACGATTAACAGGTACTTTGTTAAGATGAGAAACAGATACCGTCTTTAACAACGA 1066  
OY 488 TAAAGGAGAGTTTCAACCTTTTACCATTAATATGCAAGTCT 534  
DB 1067 ATGAATGCAAGCTATTATGGGTTTCCAAAAGATTTGTATTCT 1113

## RESULT 12

US-08-305-764C-55  
Sequence 55, Application US/08305764C  
Patent No. 5856090

GENERAL INFORMATION:

APPLICANT: Epstein, David M.

TITLE OF INVENTION: DNA METHYLASE LINKING REACTION

NUMBER OF SEQUENCES: 71

CORRESPONDENCE ADDRESS:

ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE

STREET: 10550 No. 5856090th Torrey Pines Road

CITY: La Jolla

STATE: California

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/305,764C

FILING DATE: 09-SEP-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Fitting, Thomas

REGISTRATION NUMBER: 34,163

REFERENCE/DOCKET NUMBER: TSRI 440.0

TELEPHONE: (619) 784-2937

TELEFAX: (619) 784-9399

INFORMATION FOR SEQ ID NO: 55:

SEQUENCE CHARACTERISTICS:

LENGTH: 2095 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ANTI-SENSE: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 9..1952

US-08-305-764C-55

Query Match 6.4%; Score 39.8; DB 2; Length 2095;  
Best Local Similarity 48.5%; Pred. No. 0.36;  
Matches 110; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

OY 308 GGATTCATTACGTGACAAAGTTGGGACATTATGAGCTACAAAGTTTAAATTAAT 367  
DB 887 GCATTCAGACATTACAAAAGATTATCTTTAAAGATGATGTAACCTTCT 946  
OY 368 GGAGTAATTTAATTAATAAAGGTTGAAAGATGTTGCTTCAACAAACCTTG 427  
DB 947 TAATTGACAAAATACGACTGGGACGTTAAACATTGTTCTACCTATCAAAAATTC 1006  
OY 428 TTCGACTTTACGTAGTTAGTTCAAGTAATGTTATAGGATGTTCCATATATGATTA 487  
DB 1007 AACGATTAACAGGTACTTTGTTAAGATGAGAAACAGATACCGTCTTTAACAACGA 1066  
OY 488 TAAAGGAGAGTTTCAACCTTTTACCATTAATATGCAAGTCT 534  
DB 1067 ATGAATGCAAGCTATTATGGGTTTCCAAAAGATTTGTATTCT 1113

Db 1582 GCATTGAGGACATTGCAAAAAGTATCTTTAAAAAGATGATGTAACCTTCTT 1641  
QY 368 GGAATAATTTAATATATATAAAGGTTGAAGAGATTTGCTTCAACAAACTTG 427  
Db 1642 TAATGTGCAAAAANAGACATCGGSCAGTTAAACATTAAGTTCTACTATACAAAATTC 1701  
QY 428 TTGACATTTACGTAAGTTAGTTCAAGTAAATGTTATAGTATTTGTCATATATGATTA 487  
Db 1702 AACGATTAACAGAGTACTTTTGTATAGATGAGAAACAGGATCCGCTTTTAAACAGCA 1761  
QY 488 TAAAGAGAGTTTCAAACTTTTACCATATATTTATGCAAGTCTT 534  
Db 1762 ATGAATGCAAGCTATTATAGGTTTCCAAAAGATTTGTTATTCCT 1808

## RESULT 13

US-08-323-170B-1/C  
Sequence 1, Application US/08323170B  
Patent No. 5733772  
GENERAL INFORMATION:  
APPLICANT: Williamson, Kim C.  
APPLICANT: Kaslow, David C.  
TITLE OF INVENTION: Cloning and Expression of Plasmodium  
TITLE OF INVENTION: falciparum Transmission-Blocking Target Antigen, Pf230  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/323,170B  
FILING DATE: 13-OCT-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/010,409  
FILING DATE: 29-JAN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Quine, Jonathan A.  
REGISTRATION NUMBER: P-41,261  
REFERENCE/DOCKET NUMBER: 015280-113100US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9636 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 149..9556  
US-08-323-170B-1

Query Match 6.4%; Score 39.8; DB 1; Length 9636;

Best Local Similarity 44.4%; Pred. No. 0.58; Mismatches 202; Indels 0; Gaps 0;

Matches 161; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

QY 143 CAGGCGTAATGAGGACGTCATTTTGAACAATTTCTACTATATTTTACAGAAATCATTTG 202  
Db 4973 CATCTGATCTGCGGCTTTTCTTACCTAATATGAAATTAATCTTAAATTAATAC 4914  
QY 203 GGTGAGGTTACGACACAGCTAGTATATGCTTATTTATTTATTTGCTCAAGAGATAG 262

Db 4913 GTTATGAGATTAATGTAATCCATATATGTCATAAAGATTTATTTTGTATCAGTCCAG 4854  
QY 263 CTAATTTATGATGCCAATATCTTTTATTTATGCTGTACGAGGATTTCAATACGTG 322  
Db 4853 TAACATATATATCAAAATTTATATTTATTTTATTAAGACAGCAAGAACTTAGTC 4794  
QY 323 TACAGTGTGTCATTAATGACATCAAAAGTTTAAATTTAAATGAGATTAATTTATA 382  
Db 4793 CTAAATGATATTTGACATTTATATTTTATTTTATTAATCAACATATACATTTAATG 4734  
QY 383 TAATTAAGGTTGAAGAGATGTTGCTTCTCAACAAACATTTGCACTTTACGTA 442  
Db 4733 ATATATCATTTATCCATTAAGCTGTATTTCCATATATTAATTAATGATGTTATATGTA 4674  
QY 443 GTTATGTCAGTAATTTGTTATGATGTTGTCATATATATATTAATTAAGAGAGTTT 502  
Db 4673 AGTTAGCCAAATTTATTTTAAATTAAGTATACCATTAATATTTGTTAAATTTGTATGTA 4614  
QY 503 CAA 505  
Db 4613 CTA 4611

## RESULT 14

US-08-954-441-1/C  
Sequence 1, Application US/08954441  
Patent No. 6316000  
GENERAL INFORMATION:  
APPLICANT: Williamson, Kim C.  
APPLICANT: Kaslow, David C.  
TITLE OF INVENTION: Cloning and Expression of Plasmodium  
TITLE OF INVENTION: falciparum Transmission-Blocking Target Antigen, Pf230  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/954,441  
FILING DATE: 20-OCT-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/323,170  
FILING DATE: 13-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/010,409  
FILING DATE: 29-JAN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Einhorn, Gregory P.  
REGISTRATION NUMBER: 38,440  
REFERENCE/DOCKET NUMBER: 015280-113110US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9636 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 149..9556  
US-08-954-441-1

Query Match 6.4%; Score 39.8; DB 4; Length 9636;  
Best Local Similarity 44.4%; Pred. No. 0.58;  
Matches 161; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

QY 143 CAGGGGTAATGGAGGTGCTATTTGAAACAAATTTCTACTATTTTTCACAGAAATCATTTTC 202  
DB 4973 CATCTGATCTGCTGGTGTGTTTTTTTACCTAAATCAAGAAATATTCCTTTAAATATTAAC 4914  
QY 203 GGTTCAGGTTCACACACAGTCAGTATATGCTTTATTTATTTATGTCGCAAGATATG 262  
DB 4913 GTTATGAGATTAATAGTAATCATATATGTCATAGAAATTTATTTGATAGCGCCAG 4854  
QY 263 CTATTTTATGATGCGCAATATTTACTTTTATGCTGTTACGCGATGATTTTCATTAGCTG 322  
DB 4853 TAACTATATTTCAAAATTTTATTTATTTTAAAGACAGTAAACCTTTAGGTC 4794  
QY 323 TACAGTGTGTCATATGACTACAAAGCTTTTAAATTTAAATGAGTAAATTTATA 382  
DB 4793 CTATGCGATTTGACATTTATTTTCTTTTAAATCAACATATACATTTAAATG 4734  
QY 383 TAAATAAAGGTTGAAGAAATGTTGCTTCAACAAACCTGTTGACTTTTACGTA 442  
DB 4733 ATATTCATTTATCCATTAAGCTGGATTTCCATATTAATGATTTATATGTA 4674  
QY 443 GTTTCGTCAGTATTTGTTATAGTATTTGCCATATATGATTTAAAGAGAGCTTTT 502  
DB 4673 AGTTAGCCAAATTTATTTTAAATAAAGTATACATATATTTGTTAAATTTGATGTA 4614  
QY 503 CAA 505  
DB 4613 CTA 4611

RESULT 15  
US-07-855-412B-4  
; Sequence 4, Application US/07855412B  
; Patent No. 5378819  
; GENERAL INFORMATION:  
; APPLICANT: Ryan, C.A.; McGurl, B.F.; Pearce, G.L.  
; TITLE OF INVENTION: "SYSTEMIN"  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Harness, Dickey & Pierce  
; STREET: P.O. Box 828  
; CITY: Troy  
; STATE: Michigan  
; COUNTRY: USA  
; ZIP: 48303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette-5.25 inch, 1.2mb storage  
; COMPUTER: IBM PC/386 Compatible  
; OPERATING SYSTEM: MS-DOS 4.01  
; SOFTWARE: Word for Windows-t  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/855,412B  
; FILING DATE: 03/19/92  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Smith, Deann F.  
; REGISTRATION NUMBER: 36,683  
; REFERENCE/DOCKET NUMBER: 755500002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (810) 641-1600  
; TELEFAX: (810) 641-0270  
; TELEX: 287637 Harness UR  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4526 bases  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA

DESCRIPTION: prosystemin genomic DNA, Figures 8A-8C.  
US-07-855-412B-4  
Query Match 6.3%; Score 39; DB 1; Length 4526;  
Best Local Similarity 52.8%; Pred. No. 0.75;  
Matches 84; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 196 TCATTTGCGTTGAGGTTACAGACAGTCAGTATATGCTTTATTTATTTATGTTCTCA 255  
DB 2148 TAAATTTATTTTATTTTTCAGATTTTATAGTAAAGGATTTTATTTTCAAAAA 2207  
QY 256 GAGATAGCTATTTATGATGCGCAATATTTACTTTTATGCTGTACGCGATGATTTCA 315  
DB 2208 ATGGAATCATTTTCAGAAATTAATTTATTTTGGTAACCTTTAATCTGATATATAT 2267  
QY 316 TTACGTGACAGTTGGTCATTTATGACCTCAAGGTT 354  
DB 2268 TCTCAGAGATGATGATGACCAAGAAAACTAAAGTT 2306

RESULT 16  
US-08-308-887A-4  
; Sequence 4, Application US/08308887A  
; Patent No. 5883076  
; GENERAL INFORMATION:  
; APPLICANT: Ryan, C.A.  
; APPLICANT: McGurl, B.F.  
; APPLICANT: Pearce, G.L.  
; TITLE OF INVENTION: "SYSTEMIN"  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.  
; STREET: P.O. Box 828  
; CITY: Troy  
; STATE: Michigan  
; COUNTRY: USA  
; ZIP: 48303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette-5.25 inch, 1.2mb storage  
; COMPUTER: IBM PC/386 Compatible  
; OPERATING SYSTEM: MS-DOS 4.01  
; SOFTWARE: Word for Windows-t  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/308,887A  
; FILING DATE: September 19, 1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Smith, Deann F.  
; REGISTRATION NUMBER: 36,683  
; REFERENCE/DOCKET NUMBER: 755500001USA  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (810) 641-1600  
; TELEFAX: (810) 641-0270  
; TELEX: 287637 Harness UR  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4526 bases  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
; DESCRIPTION: prosystemin genomic DNA, Figures 8A-8C.

US-08-308-887A-4  
Query Match 6.3%; Score 39; DB 2; Length 4526;  
Best Local Similarity 52.8%; Pred. No. 0.75;  
Matches 84; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 196 TCATTTGCGTTGAGGTTACAGACAGTCAGTATATGCTTTATTTATTTATGTTCTCA 255  
DB 2148 TAAATTTATTTTATTTTTCAGATTTTATAGTAAAGGATTTTATTTTCAAAAA 2207  
QY 256 GAGATAGCTATTTATGATGCGCAATATTTACTTTTATGCTGTACGCGATGATTTCA 315

Db 2208 ATGTGAATCATTTTCAGAGTTAATATTTTGGTAACCTTATCTGTATATATAT 2267  
Qy 316 TTACGTGACAGTTGGTGCAATTATGACCTCAAGGTT 354  
Db 2268 TCTCCAGAGATGATGATGACACAGAAAACCTAAGGTT 2306

## RESULT 17

US-08-881-094-4  
; Sequence 4, Application US/08881094A  
; Patent No. 6022739  
; GENERAL INFORMATION:  
; APPLICANT: Ryan, Clarence A  
; APPLICANT: Pearce, Gregory L  
; APPLICANT: McGurk, Barry F  
; TITLE OF INVENTION: Systemin  
; FILE REFERENCE: 7555-00001CPB  
; CURRENT FILING DATE: 1997-07-09  
; EARLIER FILING DATE: 1994-09-19  
; EARLIER FILING DATE: 1993-03-18  
; EARLIER FILING DATE: 1992-03-19  
; EARLIER FILING DATE: 1990-05-25  
; EARLIER FILING DATE: 1991-05-24  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 4526  
; TYPE: DNA  
; ORGANISM: Lycopersicon esculentum  
US-08-881-094-4

Query Match 6.3%; Score 39; DB 3; Length 4526;  
Best Local Similarity 52.8%; Pred. No. 0.75; Indels 0; Gaps 0;  
Matches 84; Conservative 0; Mismatches 75;

Qy 196 TCATTTCGTTGAGTTACAGCAGTCAATATGCTTATTTATTTATGCTCA 255  
Db 2148 TAAATTTATTTTATTTTTCAGATTTTATAGAGGATTTTATTTTTCAAAA 2207  
Qy 256 GAGATGATTTTATTTTATGCAATATTTATTTTATGCTTTAGGCAATGTTCA 315  
Db 2208 ATGTGAATCATTTTCAGAGTTAATATTTTGGTAACCTTATCTGTATATATAT 2267  
Qy 316 TTACGTGACAGTTGGTGCAATTATGACCTCAAGGTT 354  
Db 2268 TCTCCAGAGATGATGATGACACAGAAAACCTAAGGTT 2306

## RESULT 18

US-08-239-953-1  
; Sequence 1, Application US/08299953  
; Patent No. 5646333  
; GENERAL INFORMATION:  
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnuur  
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the  
; TITLE OF INVENTION: Expression of Foreign Proteins to the Plant Epidermis  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Woodcock Washburn Kurtz Mackiewicz & No. 5646333x1s  
; STREET: One Liberty Place 46th. Floor  
; CITY: Philadelphia  
; STATE: PA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Ver. 2.0

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/299,953  
FILING DATE: Herewith  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Beardell, Lori Y.  
REGISTRATION NUMBER: 34,293  
REFERENCE/DOCKET NUMBER: NOVA-0003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-564-8960  
TELEFAX: 215-568-3439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2861 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-239-953-1

Query Match 6.2%; Score 38.8; DB 1; Length 2861;  
Best Local Similarity 44.7%; Pred. No. 0.74; Indels 0; Gaps 0;  
Matches 151; Conservative 0; Mismatches 187;

Qy 175 TTCTACTATATTTTACAGATCAATTCGTTGAGTTACAGCAGTCAATATGCT 234  
Db 1108 TTTTACATGTTAAAGAAATTAATATGTTTATTAATACCTGAAATTTATAT 1167  
Qy 235 TTAATTTATGTTGTCAGAGATGCTATTTATGATGCCAATTTATTTAT 294  
Db 1168 TATCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1227  
Qy 295 GCTGTACGAGATGATTTCACTTACGTGATGATGATGATGATGATGATGATGAT 354  
Db 1228 TTTTGAATATATTTTATTAATGATGATGATGATGATGATGATGATGATGAT 1287  
Qy 355 TTTAATTTAATGAGTAAATTTAATTAATTAATTAATTAATTAATTAATTAAT 414  
Db 1288 TTTTAAAGTCAATTTTATGATTTTTCACATATCAAGAGATCAATATTTAT 1347  
Qy 415 CAAGAAACATGTTTCAGCTTTTACGTATGATGATGATGATGATGATGATGAT 474  
Db 1348 TTTCAAAATTAATTTCTTTTATTAATTAATTAATTAATTAATTAATTAAT 1407  
Qy 475 CCATATATGATTAATTAAGAGAGATTTTCAACTTTT 512  
Db 1408 CTTTTCCTCTCCATCAATTAATTAATTAATTAATTAATTAATTAATTAAT 1445

## RESULT 19

US-08-459-415-1  
; Sequence 1, Application US/08459415  
; Patent No. 5744334  
; GENERAL INFORMATION:  
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnuur  
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the  
; TITLE OF INVENTION: Expression of Foreign Proteins to the Plant Epidermis  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Woodcock Washburn Kurtz Mackiewicz & No. 5744334x1s  
; STREET: One Liberty Place 46th. Floor  
; CITY: Philadelphia  
; STATE: PA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25

SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,415  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/299,953  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Beardell, Lori Y.  
REGISTRATION NUMBER: 34,293  
REFERENCE/DOCKET NUMBER: NOVA-0003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-564-8960  
TELEFAX: 215-568-3439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2861 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-459-415-1

Query Match 6.2%; Score 38.8; DB 1; Length 2861;  
Best Local Similarity 44.7%; Pred. No. 0.74;  
Matches 151; Conservative 0; Mismatches 187; Indels 0; Gaps 0;

175 TTCTACTATATTTTACAGATCATTTGGTTGAGGTACAGACAGTCAGTAAATGCT 234  
1108 TTTTCAATGTTAAAGAAATTAATTAATTTTATTAATTAATTAATTAATTAAT 1167  
235 TTATTTATTTAGTGTGCTCAAGATAGCTATTTATGAGCAATTAATTAATTT 294  
1168 TATCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1227  
295 GCTGTACGCGATGATTTCACTGATGACGTAAGGTTGAGCTTAAAGCAAGTT 354  
1228 TTTTGAATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 1287  
355 TTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 414  
1288 TTTTAAAGTGTATTTTGTGATTTTGGACATATCAAGAGGATATTAATTTGTT 1347  
415 CAACAACACTGTGCTGATTTTACGATTTTACGATTTTACGATTTTACGATTT 474  
1348 TTTCAACATTAATTTCTTTTACTTAATTAATTAATTAATTAATTAATTAAT 1407  
475 CCATATATGATTTAAAGAGAGTTTCAACTTTT 512  
1408 CTTTCTCTCTCTCCATCATTAATTAATTAATTAATTAATTAATTAATTAAT 1445

RESULT 20  
US-09-066-687-1  
Sequence 1, Application US/09066687  
Patent No. 6339185  
GENERAL INFORMATION:  
APPLICANT: Dobres, Michael S. and Mandaci, Sevnuur  
TITLE OF INVENTION: A Plant Promoter Useful for Directing the  
TITLE OF INVENTION: Expression of Foreign Proteins to the Plant Epidermis  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6339185crls  
STREET: One Liberty Place 46th. Floor  
CITY: Philadelphia  
STATE: PA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/066,687  
FILING DATE: Herewith  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Beardell, Lori Y.  
REGISTRATION NUMBER: 34,293  
REFERENCE/DOCKET NUMBER: NOVA-0003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-564-8960  
TELEFAX: 215-568-3439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2861 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-09-066-687-1

Query Match 6.2%; Score 38.8; DB 4; Length 2861;  
Best Local Similarity 44.7%; Pred. No. 0.74;  
Matches 151; Conservative 0; Mismatches 187; Indels 0; Gaps 0;

175 TTCTACTATATTTTACAGATCATTTGGTTGAGGTACAGACAGTCAGTAAATGCT 234  
1108 TTTTCAATGTTAAAGAAATTAATTAATTTTATTAATTAATTAATTAATTAAT 1167  
235 TTATTTATTTAGTGTGCTCAAGATAGCTATTTATGAGCAATTAATTAATTT 294  
1168 TATCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1227  
295 GCTGTACGCGATGATTTCACTGATGACGTAAGGTTGAGCTTAAAGCAAGTT 354  
1228 TTTTGAATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 1287  
355 TTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 414  
1288 TTTTAAAGTGTATTTTGTGATTTTGGACATATCAAGAGGATATTAATTTGTT 1347  
415 CAACAACACTGTGCTGATTTTACGATTTTACGATTTTACGATTTTACGATTT 474  
1348 TTTCAACATTAATTTCTTTTACTTAATTAATTAATTAATTAATTAATTAAT 1407  
475 CCATATATGATTTAAAGAGAGTTTCAACTTTT 512  
1408 CTTTCTCTCTCTCCATCATTAATTAATTAATTAATTAATTAATTAATTAAT 1445

RESULT 21  
PCT-US95-11231-1  
Sequence 1, Application PCT/US9511231  
GENERAL INFORMATION:  
APPLICANT: Dobres, Michael S. and Mandaci, Sevnuur  
TITLE OF INVENTION: A Plant Promoter Useful for Directing the Expression  
TITLE OF INVENTION: of Foreign Proteins to the Plant Epidermis  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & Norris  
STREET: One Liberty Place 46th. Floor  
CITY: Philadelphia  
STATE: PA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/11231

FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/299,953  
FILING DATE: September 2, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Beardell, Lori Y.  
REGISTRATION NUMBER: 34,293  
REFERENCE/DOCKET NUMBER: NOVA-0016  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-564-8960  
TELEFAX: 215-568-3439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2861 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
PCT-US95-11231-1

Query Match 6.2%; Score 38.8; DB 5; Length 2861;  
Best Local Similarity 44.7%; Pred. No. 0.74;  
Matches 151; Conservative 0; Mismatches 187; Indels 0; Gaps 0;

Qy 175 TTCTACTATATTTTACAGATCATTCGTTGAGTTACAGACAGTCAATATGCT 234  
Db 1108 TTTTACATGTTAAAGAAATTAATTAATTTTATTAATTAATTAATTAAT 1167  
Qy 235 TTATTTATTTATGTTGTCACAGATAGCTATTTATGATGCAATATTAATTTTAT 294  
Db 1168 TATCTTTTATTAATTAATTAATTAATTTTATTAATTAATTAATTAATTAAT 1227  
Qy 295 GCTGTACGATGATTCATTAAGTACAGATGTTGATTAATGATTAATTAATTAAT 354  
Db 1228 TTTTGTATTAATTTTATTAATGATTAATTAATTAATTAATTAATTAATTAAT 1287  
Qy 355 TTTAATTTAATGAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 414  
Db 1288 TTTTAAGTCATTTTATTAATTTTATTAATTAATTAATTAATTAATTAATTAAT 1347  
Qy 415 CACAAACACTGTTGACATTTTACGTTAGTTCAAGTAATTTGTAATGTAATGTT 474  
Db 1348 TTTCAACATTAATTTCTTTTACTTAATTAATTAATTAATTAATTAATTAATTAAT 1407  
Qy 475 CCATATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 512  
Db 1408 CTTTTCCTCTCTCCATCATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1445

RESULT 22  
US-08-299-953-2  
Sequence 2, Application US/08299953  
Patent No. 5646333

GENERAL INFORMATION:  
APPLICANT: Dobres, Michael S. and Mandaci, Sevnur  
TITLE OF INVENTION: A Plant Promoter Useful for Directing the  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESS: Woodcock Washburn Kurtz Mackiewicz & No. 5646333r1s  
STREET: One Liberty Place 46th. Floor  
CITY: Philadelphia  
STATE: PA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/299,953  
FILING DATE: Herewith  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Beardell, Lori Y.  
REGISTRATION NUMBER: 34,293  
REFERENCE/DOCKET NUMBER: NOVA-0003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-564-8960  
TELEFAX: 215-568-3439  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3881 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-299-953-2

Query Match 6.2%; Score 38.8; DB 1; Length 3881;  
Best Local Similarity 44.7%; Pred. No. 0.81;  
Matches 151; Conservative 0; Mismatches 187; Indels 0; Gaps 0;

Qy 175 TTCTACTATATTTTACAGATCATTCGTTGAGTTACAGACAGTCAATATGCT 234  
Db 1108 TTTTACATGTTAAAGAAATTAATTAATTTTATTAATTAATTAATTAATTAAT 1167  
Qy 235 TTATTTATTTATGTTGTCACAGATAGCTATTTATGATGCAATATTAATTTTAT 294  
Db 1168 TATCTTTTATTAATTAATTAATTAATTTTATTAATTAATTAATTAATTAAT 1227  
Qy 295 GCTGTACGATGATTCATTAAGTACAGATGTTGATTAATGATTAATTAATTAAT 354  
Db 1228 TTTTGTATTAATTTTATTAATGATTAATTAATTAATTAATTAATTAATTAAT 1287  
Qy 355 TTTAATTTAATGAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 414  
Db 1288 TTTTAAGTCATTTTATTAATTTTATTAATTAATTAATTAATTAATTAATTAAT 1347  
Qy 415 CACAAACACTGTTGACATTTTACGTTAGTTCAAGTAATTTGTAATGTAATGTT 474  
Db 1348 TTTCAACATTAATTTCTTTTACTTAATTAATTAATTAATTAATTAATTAATTAAT 1407  
Qy 475 CCATATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 512  
Db 1408 CTTTTCCTCTCTCCATCATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1445

RESULT 23  
US-08-459-415-2  
Sequence 2, Application US/08459415  
Patent No. 5744334

GENERAL INFORMATION:  
APPLICANT: Dobres, Michael S. and Mandaci, Sevnur  
TITLE OF INVENTION: A Plant Promoter Useful for Directing the  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESS: Woodcock Washburn Kurtz Mackiewicz & No. 5744334r1s  
STREET: One Liberty Place 46th. Floor  
CITY: Philadelphia  
STATE: PA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,415  
FILING DATE: 02-JUN-1995



CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/299,953  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Beardell, Lori Y.  
REGISTRATION NUMBER: 34,293  
REFERENCE/DOCKET NUMBER: NOVA-0003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-564-8960  
TELEFAX: 215-564-3439  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3881 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-459-415-2

Query Match 6.2%; Score 38.8; DB 1; Length 3881;  
Best Local Similarity 44.7%; Pred. No. 0.81;  
Matches 151; Conservative 0; Mismatches 187; Indels 0; Gaps 0;

QY 175 TTCTACTATATTTTACAGATCATTCGGTTGAGGTTACGACAGTCAGTATATGCT 234  
DB 1108 TTTTACATGTTAAAGAAATTAATTTATTTATTAATTAATTAATTAATTAAT 1167  
QY 235 TTATTTATTTATGTTGCTCAAGATAGCTATTTATTTATTTATTTATTTATTT 294  
DB 1168 TATCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1227  
QY 295 GCTGTACGAGATGATTTTCAATTTACGTATCAAGTGTGATTTGATTTGATTTGAT 354  
DB 1228 TTTTGAATATATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1287  
QY 355 TTTTAAATTTAAATGAGTAATTTATTTATTTATTTATTTATTTATTTATTTATTT 414  
DB 1288 TTTTAAAGTCATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1347  
QY 415 CAACAACACTGTTGACCTTTTACGTATTTAGTTCAAGTAATTTGTAATTTGTT 474  
DB 1348 TTTTCAACAAATATTTCTTTTACTTTAATTAATCCCTTAATTTATTTATTTATTT 1407  
QY 475 CCATATATGATTTATTAAGAGAGTTTCAACTTTT 512  
DB 1408 CTTTCTCTCTCTCCAAATCATTTATCTAGATAATTT 1445

RESULT 24  
US-09-066-687-2  
Sequence 2, Application US/0906687  
Patent No. 6339185  
GENERAL INFORMATION:  
APPLICANT: Dobres, Michael S. and Mandaci, Sevnur  
TITLE OF INVENTION: A Plant Promoter Useful for Directing the  
TITLE OF INVENTION: Expression of Foreign Proteins to the Plant Epidermis  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6339185r1s  
STREET: One Liberty Place 46th. Floor  
CITY: Philadelphia  
STATE: PA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/066,687

FILING DATE: Herewith  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Beardell, Lori Y.  
REGISTRATION NUMBER: 34,293  
REFERENCE/DOCKET NUMBER: NOVA-0003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-564-8960  
TELEFAX: 215-564-3439  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3881 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-09-066-687-2

Query Match 6.2%; Score 38.8; DB 4; Length 3881;  
Best Local Similarity 44.7%; Pred. No. 0.81;  
Matches 151; Conservative 0; Mismatches 187; Indels 0; Gaps 0;

QY 175 TTCTACTATATTTTACAGATCATTCGGTTGAGGTTACGACAGTCAGTATATGCT 234  
DB 1108 TTTTACATGTTAAAGAAATTAATTTATTTATTTATTTATTTATTTATTTATTTAT 1167  
QY 235 TTATTTATTTATGTTGCTCAAGATAGCTATTTATTTATTTATTTATTTATTTATTT 294  
DB 1168 TATCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1227  
QY 295 GCTGTACGAGATGATTTTCAATTTACGTATCAAGTGTGATTTGATTTGATTTGAT 354  
DB 1228 TTTTGAATATATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1287  
QY 355 TTTTAAATTTAAATGAGTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 414  
DB 1288 TTTTAAAGTCATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1347  
QY 415 CAACAACACTGTTGACCTTTTACGTATTTAGTTCAAGTAATTTGTAATTTGTT 474  
DB 1348 TTTTCAACAAATATTTCTTTTACTTTAATTAATCCCTTAATTTATTTATTTATTTATTT 1407  
QY 475 CCATATATGATTTATTAAGAGAGTTTCAACTTTT 512  
DB 1408 CTTTCTCTCTCTCCAAATCATTTATCTAGATAATTT 1445

RESULT 25  
PCT-US95-11231-2  
Sequence 2, Application PC/TUS9511231  
GENERAL INFORMATION:  
APPLICANT: Dobres, Michael S. and Mandaci, Sevnur  
TITLE OF INVENTION: A Plant Promoter Useful for Directing the Expression  
TITLE OF INVENTION: of Foreign Proteins to the Plant Epidermis  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & Norris  
STREET: One Liberty Place 46th. Floor  
CITY: Philadelphia  
STATE: PA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/11231  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/299,953  
FILING DATE: September 2, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Beardsell, Lori Y.  
REGISTRATION NUMBER: 34,293  
REFERENCE/DOCKET NUMBER: NOVA-0016  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-564-8960  
TELEFAX: 215-568-3439  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3881 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
PCT-US95-11231-2

Query Match 6.2%; Score 38.8; DB 5; Length 3881;  
Best Local Similarity 44.7%; Pred. No. 0.81; Mismatches 187; Indels 0; Gaps 0;  
Matches 151; Conservative 0; Mismatches 187; Indels 0; Gaps 0;

QY 175 TTCTACTATATTTTACAGATTCCTCGTTGAGTTACAGACAGTCAATATGCT 234  
DB 1108 TTTTACATGTTAAAGATTAATATGTTTATATATACCTAGAAATTTATATAT 1167  
QY 235 TTATTTATTTATGTCGTCAGAGATAGCTATTTATGATGCCAATTTACTTTTAT 294  
DB 1168 TATCTTTTATTTATATATATATATATATATATATATATATATATATATAT 1227  
QY 295 GCTGTACGCGATGATTTCTATTCGTCATCAAGTGGTCATATGACTACAAAGTT 354  
DB 1228 TTTTGTATATATTTTATATAGCTGTAATAATCTAATAATTTACTCCCTCGTTCT 1287  
QY 355 TTTAATTTAATGAGTAAATTTAATATATATATATATATATATATATATATAT 414  
DB 1288 TTTTAAAGTGCATTTTATGATTTTGGACATATCAAGAGATATATATATATAT 1347  
QY 415 CAACAAACACTGTTGCACTTTTACGTAATTTAGTTCAAGTAATTTGTAATGTA 474  
DB 1348 TTTTAAACAAATTAATCTCTCTTTTACTATATATATATATATATATATATAT 1407  
QY 475 CCATATATGATTATATATATATATATATATATATATATATATATATATAT 512  
DB 1408 CTTTCTCTCTCTCCATATCTATCTAGAGATATATAT 1445

## RESULT 26

US-08-916-421B-1  
Sequence 1, Application US/08916421B  
Patent No. 6503729  
GENERAL INFORMATION:  
APPLICANT: Bull et al.  
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus  
Patent No. 6503729  
TITLE OF INVENTION: jannaschii  
FILE REFERENCE: PB275  
CURRENT APPLICATION NUMBER: US/08/916,421B  
CURRENT FILING DATE: 1997-08-22  
PRIOR APPLICATION NUMBER: US 60/024,428  
PRIOR FILING DATE: 1996-08-22  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 1664976  
TYPE: DNA  
ORGANISM: Methanococcus jannaschii  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (28222)..(28222)  
OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc feature  
LOCATION: (28257)..(28258)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (84773)..(84773)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (84808)..(84808)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (84812)..(84812)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (98120)..(98120)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (98159)..(98159)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (98239)..(98239)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (98266)..(98266)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (103998)..(103998)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (148948)..(148948)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (163385)..(163385)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (191989)..(191989)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (191995)..(191995)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (231980)..(231980)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (234187)..(234187)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (234220)..(234220)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (234814)..(234814)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (309398)..(309398)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (309418)..(309418)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (312837)..(312837)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (312993)..(312993)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (319226)..(319226)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (559167)..(559167)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature

LOCATION: (559241)..(559241)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (600992)..(600992)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (622708)..(622708)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (657081)..(657081)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (657203)..(657203)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (674435)..(674435)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (682442)..(682442)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (713652)..(713652)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (741684)..(741684)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (779455)..(779455)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (779676)..(779676)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (855539)..(855539)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (871619)..(871619)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1084830)..(1084830)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1096846)..(1096846)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1119881)..(1119881)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1130881)..(1130881)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1310988)..(1310988)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1313224)..(1313224)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1349473)..(1349473)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1349491)..(1349491)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1470091)..(1470091)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1569020)..(1569020)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1602912)..(1602912)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1603734)..(1603734)

OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1637998)..(1637998)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1664854)..(1664854)  
OTHER INFORMATION: n equals a, t, c, or g  
US-08-916-421B-1  
Query Match 6.2%; Score 38.6; DB 4; Length 1664976;  
Best Local Similarly 48.4%; Pred. No. 5;  
Matches 107; Conservative 0; Mismatches 114; Indels 0; Gaps 0;  
QY 323 TACAGTTGGTCATGCTTATGCTCAAGCTTTTAAATGAGTAAATTTATA 382  
DB 1179668 TAGAAGTTGATGAGGCAAGATTAAGACCTTTAAAGATGGAAGCTAAATACATG 1179727  
QY 383 TAATMAAGGCTTGAAGAAATGTTGCTTCTCAACAACACTTGTTCAGCTTTACGTA 442  
DB 1179728 ATGAAGAAGTCATTAAGGAATATTTTACAGACCACTTAATTTTAAAGATGCTGTTA 1179787  
QY 443 GTTAGCTCAAGTAATGTTATAGTATGTTCATATATGATTAAGAGAGTTT 502  
DB 1179788 AAAAGCTTGAATTAATTTATTTGCTTTTATTAATACCTCATTTGCAATTT 1179847  
QY 503 CAACTTTTACATTAATATGCAAGCTTCAGGTG 543  
DB 1179848 TACATTTTACATTTATTTATTTTAAATTTTAAATATG 1179888  
RESULT 27  
US-08-998-416-701  
Sequence 701, Application US/08998416  
Patent No. 6239264  
GENERAL INFORMATION:  
APPLICANT: Philippsen, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jürgen  
APPLICANT: Knechtle, Philipp  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPIL  
TITLE OF INVENTION: AND USES THEREOF  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6239264artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998, 416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38, 241  
REFERENCE/DOCKET NUMBER: PF/5-30306/N/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 701:  
SEQUENCE CHARACTERISTICS:

LENGTH: 701 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAG1462RP  
US-08-998-416-701

Query Match 6.1%; Score 37.8; DB 3; Length 701;  
Best Local Similarity 48.0%; Pred. No. 0.89;  
Matches 108; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 157 CGTCATTTGGAACAATTTTCTACTATTTTTCAGATCATTTGCGTTGAGTTCA 216  
DB 475 CTTTATTTGATATTTACTACTACTAATATTTACTAATATATATATTAAGAACT 534  
QY 217 GCACAGTCAGTATATGCTTATTTATTTATGTTGCTCAAGATAGTATTTTATGATG 276  
DB 535 TAAATCTAATATTTATTTATCTAAAGTATATATAATTAATTAACCTTTTATTTATTTT 594  
QY 277 CCAATATTACTTTTATTTGCTGTTACGCGCATGATTCATTAACGTACAAAGTTGTCGA 336  
DB 595 AATATTTATTTATTTAGTAAATTTATTTATTTATTTATTTATTAACATTAATTTTGGATGA 654  
QY 337 TTATGACTACAAAGTTTAAATTTAAATGAGTAAATTTAAT 381  
DB 655 TAAATATCATTTAATTAAGTAAATTTATTAATAATTTATCTTAAT 699

RESULT 28  
US-09-134-001C-595  
Sequence 595, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 595  
LENGTH: 1401  
TYPE: DNA  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-595

Query Match 6.1%; Score 37.8; DB 4; Length 1401;  
Best Local Similarity 48.0%; Pred. No. 1.1;  
Matches 108; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 280 ATATTACTTTTATTTGCTGTTACGCGCATGATTCATTAGCGTACAAAGTTGGCATTA 339  
DB 841 ATACAGCTATCATCGCTTTTGTAGTCAATTCCTTATTTACTTACCTTGGTGCAGCACTA 900  
QY 340 TGACACTCAAAAGTTTTTAAATTTAAATGAGTAAATTTAATATATATAAAGGGTTGAA 399  
DB 901 TTTTGTGAATAAAGCTGAAGATTTGGGTTTATATGATCTTTATTTGGCTTAA 960  
QY 400 GGAATGTTGCTTTCACAAACACTTGTGACTTTTACGTAGTTTAGTTCAAGTATTT 459  
DB 961 ACAAACTGCTTTAGGTGACACCTTGGCGGTATTTAGACTCTTTATTTGCTGGCC 1020  
QY 460 GTTATAGTATTTGCTCATATATATATATAAAGAGAGTTTCA 504  
DB 1021 CTTTATGCTTCTGTCAAATTTCAACTATTAACAGAGAGTTAGCA 1065

RESULT 29  
US-09-107-532A-905/c  
Sequence 905, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354

COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Arindello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 905:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2142 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...2142  
SEQUENCE DESCRIPTION: SEQ ID NO: 905:  
US-09-107-532A-905

Query Match 6.1%; Score 37.8; DB 4; Length 2142;  
Best Local Similarity 47.5%; Pred. No. 1.2;  
Matches 144; Conservative 0; Mismatches 157; Indels 2; Gaps 1;

QY 230 ATGCTTATTTATTTATGTTGCTCAAGAGATGCTATTTTATGATGCCAATATTACTTT 289  
DB 1630 ATGATTCGTATTTCTGATATATATCTTAACCTATGAGATCGATGATATGTAATAT 1571  
QY 290 TTATGCTGTTAGCGCATGATTCATTAGCGTACAAAGTTGTCATTAAGCACTCAA 349  
DB 1570 CTTTGGCTTCCATATAAGATTTAAACTTAATAATCTTTTGGTTATTTTAACAATA 1511  
QY 350 AGGTTTAAATTTAAATGAGTAAATTTAATATATATAAAGGTTGAAAGAAATGTTTG 409  
DB 1510 TTAGTTTATTTTAAATTTTTCAAATTTAGAGATTTATAAGAGCTTGAGA--GGAGA 1453  
QY 410 CTTCTCAACAACTGTTGACTTTTACGTAGTTTAGTTAGTTCAAGTAAATGTTATAGTA 469  
DB 1452 GTCTCAAGCTCAAACTGATATGCTCTAAATATATTTTTCATCATCTTTTGCATATTC 1393

Qy 470 TTGTTCCATATATGATTTAAAGAGAGTTTCAACTTTTACCATTATATATGCAA 529  
Db 1392 TTTATATTTTGTAGTACTAGATGTAAGTATGTCATCAGAGTTTATTAATTTAA 1333  
Qy 530 GTC 532  
Db 1332 GTC 1330

## RESULT 30

US-09-489-847-67/c  
Sequence 67, Application US/09489847  
Patent No. 6476195  
GENERAL INFORMATION:  
APPLICANT: Rosen et al  
TITLE OF INVENTION: 98 Human Secreted Proteins  
FILE REFERENCE: P2031P1  
CURRENT APPLICATION NUMBER: US/09/489,847  
EARLIER FILING DATE: 2000-01-24  
EARLIER APPLICATION NUMBER: PCT/US99/17130  
EARLIER FILING DATE: 1999-07-29  
EARLIER APPLICATION NUMBER: 60/094,657  
EARLIER FILING DATE: 1998-07-30  
EARLIER APPLICATION NUMBER: 60/095,486  
EARLIER FILING DATE: 1998-08-05  
EARLIER APPLICATION NUMBER: 60/096,319  
EARLIER FILING DATE: 1998-08-12  
EARLIER APPLICATION NUMBER: 60/095,454  
EARLIER FILING DATE: 1998-08-06  
EARLIER APPLICATION NUMBER: 60/095,455  
EARLIER FILING DATE: 1998-08-06  
NUMBER OF SEQ ID NOS: 376  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 67  
LENGTH: 2434  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (10)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (12)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (27)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (73)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (75)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (103)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (130)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-489-847-67

Query Match 6.1%; Score 37.8; DB 4; Length 2434;  
Best Local Similarity 45.5%; Pred. No. 1.3; 162; Indels 0; Gaps 0;  
Matches 135; Conservative 0; Mismatches

Qy 227 TATAGCTTATTTATTTATTTGTCACAGATAGCTATTTTATTGATGCCAATATTAC 286

Db 2421 TT 2362  
Qy 287 TTTTATATGCTGTAGGAGATTTTCATTACGTGACAGAGTTGGCATATAGACTA 346  
Db 2361 TTTTATTTTTTTTTTATGTTCTGTAATTTTAAATTTCTTACGGGTCATTTACTTA 2302  
Qy 347 CAAGGTTTTTAAATTTAAATGAGATTAATTTATATTAATAAGGGTTGAAAGGAATGT 406  
Db 2301 AAAAAATATATAAATTTTAAATGATCTTAGAAAAAATTAGCAAAAAATTGAGAAAGTT 2242  
Qy 407 TTGCTTCTCAACAACTGTTGACTTTTACGTAGTTAGTCAAGTAATGTTANAG 466  
Db 2241 GAGCATATTTACATTTGACAAATGTTTTCATTCGAAATGTTTAAATCATTAATTTGGTT 2182  
Qy 467 GTATGTTCCATATATGATTTAAAGAGAGTTTCAACTTTTACCATTATAT 523  
Db 2181 TCCCTCAACATACACATCAAAAATACATTTCAAAAACCTTTTTCATATATAT 2125

## RESULT 31

US-07-641-143B-1  
Sequence 1, Application US/07641143B  
Patent No. 5436000  
GENERAL INFORMATION:  
APPLICANT: Bardour, Alan G.  
APPLICANT: Bundoc, Virgilio  
TITLE OF INVENTION: Flagella-less Borrelia  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: TX  
COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/641,143B  
FILING DATE: 11-JAN-1991  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Patterson, Melinda  
REGISTRATION NUMBER: 33,062  
TELECOMMUNICATION INFORMATION:  
REFERENCE/DOCKET NUMBER: UTSK092  
TELEPHONE: 713-787-1592  
TELEFAX: 713-789-2679  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1507 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Borrelia burgdorferi  
STRAIN: HBI9  
IMMEDIATE SOURCE:  
LIBRARY: PAC1  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 207..1217  
OTHER INFORMATION: /product= "Flagellin"  
FEATURE:  
NAME/KEY: RBS  
LOCATION: 194..198  
FEATURE:  
NAME/KEY: 10\_signal  
LOCATION: 146..151

FEATURE:  
NAME/KEY: -35\_signal  
LOCATION: 122..125  
US-07-641-143B-1

Query Match 6.0%; Score 37.6; DB 1; Length 1507;  
Best Local Similarity 46.8%; Pred. No. 1.2;  
Matches 118; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 217 GCACAGTCAGTATGCTTTTATTATTATGCTGCAAGAGATGCTATTATTGATG 276  
DB 1083 GCTCAATTAAGAATGCTACATGACATGAGTTGTAGACGCAACTAATATGATTT 1142  
QY 277 CCAATATTACTTTTATTGCTGTACGCGATGATTCATTACGTTACAGTTGTCGA 336  
DB 1143 TTAACACATCTGCAATGCGCATGATTCGACGGCTATCAAGTCCCAATATGTTTG 1202  
QY 337 TTATGACTACAAAGCTTTTAAATTAATGAGTAAATTAATATATATAAAGGTTG 396  
DB 1203 TCATTTGCTTATGATTAATTTCTAGTTTAAATGTTAATTAATTAACAAAGATCCCTT 1262  
QY 397 AAAGAAATGTTGCTTCTCAACAACACTTGTGACTTTAGCTTATGTTCAAGTA 456  
DB 1263 AAAGATCTTTTGTTTTATTTCAGATCGCAAAATTAATTAATTTAGTAAATTT 1322  
QY 457 ATTGTTATAGT 468  
DB 1323 ATATGATGTGT 1334

## RESULT 32

US-08-124-290-1

Sequence 1, Application US/08124290

Patent No. 5585102

GENERAL INFORMATION:

APPLICANT: Barbour, Alan G.

TITLE OF INVENTION: Flagella-less Borrelia

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: TX

COUNTRY: USA

ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/124,290

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/641,143

FILING DATE: 11-JAN-1991

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Patterson, Melinda

REGISTRATION NUMBER: 33,062

REFERENCE/DOCKET NUMBER: UTSK092

TELECOMMUNICATION INFORMATION:

TELEPHONE: 713-787-1592

TELEFAX: 713-789-2679

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1507 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:  
ORGANISM: Borrelia burgdorferi  
STRAIN: HB19  
IMMEDIATE SOURCE:  
LIBRARY: PACBI

FEATURE:  
NAME/KEY: CDS  
LOCATION: 207..1217  
OTHER INFORMATION: /product="Flagellin"

FEATURE:  
NAME/KEY: RBS  
LOCATION: 194..198

FEATURE:  
NAME/KEY: 10\_signal  
LOCATION: 146..151

FEATURE:  
NAME/KEY: -35\_signal  
LOCATION: 122..125  
US-08-124-290-1

Query Match 6.0%; Score 37.6; DB 1; Length 1507;  
Best Local Similarity 46.8%; Pred. No. 1.2;  
Matches 118; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 217 GCACAGTCAGTATGCTTTTATTATTATGCTGCTCAAGATGCTATTATTGATG 276  
DB 1083 GCTCAATTAAGAATGCTACATGACATGAGTTGTACGACGCAACTAATATGATTT 1142  
QY 277 CCAATATTACTTTTATTGCTGTACGCGATGATTCATTACGTTACAGTTGTCGA 336  
DB 1143 TTAACACATCTGCAATGCGCATGATTCGACGGCTATCAAGTCCCAATATGTTTG 1202  
QY 337 TTATGACTACAAAGCTTTTAAATTAATGAGTAAATTAATATATAAAGGTTG 396  
DB 1203 TCATTTGCTTATGATTAATTTCTAGTTTAAATGTTAATTAATTAACAAAGATCCCTT 1262  
QY 397 AAAGAAATGTTGCTTCTCAACAACACTTGTGACTTTAGCTTATGTTCAAGTA 456  
DB 1263 AAAGATCTTTTGTTTTATTTCAGATCGCAAAATTAATTAATTTAGTAAATTT 1322  
QY 457 ATTGTTATAGT 468  
DB 1323 ATATGATGTGT 1334

## RESULT 33

US-08-696-372A-1

Sequence 1, Application US/08696372A

Patent No. 6077515

GENERAL INFORMATION:

APPLICANT: Barbour, Alan G.

TITLE OF INVENTION: Flagella-less Borrelia

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee

STREET: PO Box 4433

CITY: Houston

STATE: Texas

COUNTRY: USA

ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/696,372A

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/641,143

FILING DATE: 11-JAN-1991

```

Z1P: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCI
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Pamela Daneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 3402:
SEQUENCE CHARACTERISTICS:
LENGTH: 1296 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium

```

```

; ; NAME/KEY: misc feature
; ; LOCATION: (B) LOCATION 1...1296
; ; SEQUENCE DESCRIPTION: SEQ ID NO: 3402:
75 00 105 500 3402

```

[illegible]

```

QY      332  TAAATTATATTAATAAAAAAGGTGAAGAAGATGTTCTCTCAACAACTTGTTCG 433
Db      441  TAAATAATAGAGAAATAATGATTAAGAAATTCATATATGTTCTCATGAAGAATTGTAAACA 500
QY      432  ACTTTACGTAGTTTAGTTCAAGTATTTGTTATAG 466
Db      501  AACCAATCTAAAAATTAGGTGATGAAGATTGATCTTG 535

RESULT 35
US-08-286-325A-1
; Sequence 1, Application US/08286325A
; Patent No. 5658770
; GENERAL INFORMATION:
; APPLICANT: PREVOTS, Fabien

```



APPLICANT: REMY, Elisabeth  
APPLICANT: RITZENTHALER, Paul  
TITLE OF INVENTION: Nucleic acid sequence and plasmids  
TITLE OF INVENTION: comprising at least one phage resistance mechanism,  
TITLE OF INVENTION: bacteria in which they are present, and their use  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BACON & THOMAS  
STREET: 625 Slaters Lane - Fourth Floor  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22314  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/286,325A  
FILING DATE: 04-AUG-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 93 09 777  
FILING DATE: 09-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: FICHTER, Richard E  
REGISTRATION NUMBER: 26,382  
REFERENCE/DOCKET NUMBER: REF/BDL/  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 683-0500  
TELEFAX: (703) 683-1080  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1875 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Lactococcus lactis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 202..1821  
US-08-286-325A-1

Query Match 6.0%; Score 37.4; DB 1; Length 1875;  
Best Local Similarity 51.5%; Pred. No. 1.5;  
Matches 86; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 249 TGCTCAAGAGTAGCTATTATTGATGCCAATATTACTTTTATGCTGTAGCGCATG 308  
DB 75 TGTCCAGAAATGACATTTTCTGAAACCCATATTAATAATTTTGTATATCCCATATA 134  
QY 309 GATTTCATTACGTGTACAGTTGGTGCATTTGACACTACCAAGGTTTAAATTTAAATG 368  
DB 135 TATTATATATAGCTTCAATGTATAAATTTATATGATATATATAGAAATTTTAAAA 194  
QY 369 GAGTAATTTATATATATATAAAGGTTGAAGAAGATGTTTCTCTC 415  
DB 195 AATATGATGATATATATATGACTTTAAACATATGTTAAGCTATC 241

RESULT 36  
US-08-286-325A-7  
Sequence 7, Application US/08286325A  
GENERAL INFORMATION:  
APPLICANT: BRENOTS, Fabien  
APPLICANT: REMY, Elisabeth  
APPLICANT: RITZENTHALER, Paul

TITLE OF INVENTION: Nucleic acid sequence and plasmids  
TITLE OF INVENTION: comprising at least one phage resistance mechanism,  
TITLE OF INVENTION: bacteria in which they are present, and their use  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BACON & THOMAS  
STREET: 625 Slaters Lane - Fourth Floor  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22314  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/286,325A  
FILING DATE: 04-AUG-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 93 09 777  
FILING DATE: 09-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: FICHTER, Richard E  
REGISTRATION NUMBER: 26,382  
REFERENCE/DOCKET NUMBER: REF/BDL/  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 683-0500  
TELEFAX: (703) 683-1080  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3234 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Lactococcus lactis  
US-08-286-325A-7

Query Match 6.0%; Score 37.4; DB 1; Length 3234;  
Best Local Similarity 51.5%; Pred. No. 1.7;  
Matches 86; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 249 TGCTCAAGAGTAGCTATTATTGATGCCAATATTACTTTTATGCTGTAGCGCATG 308  
DB 587 TGTCCAGAAATGACATTTTCTGAAACCCATATTAATAATTTTGTATATCCCATATA 646  
QY 309 GATTTCATTACGTGTACAGTTGGTGCATTTGACACTACCAAGGTTTAAATTTAAATG 368  
DB 647 TATTATATATAGCTTCAATGTATAAATTTATATGATATATATAGAAATTTTAAAA 706  
QY 369 GAGTAATTTATATATATATAAAGGTTGAAGAAGATGTTTCTCTC 415  
DB 707 AATATGATGATATATATATGACTTTAAACATATGTTAAGCTATC 753

RESULT 37  
US-09-134-001C-2508  
Sequence 2508, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC-007  
CURRENT FILING DATE: 1998-08-13  
CURRENT APPLICATION NUMBER: US/09/134,001C  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779  
 PRIOR FILING DATE: 1997-08-14  
 NUMBER OF SEQ ID NOS: 5674  
 SEQ ID NO 2508  
 LENGTH: 576  
 TYPE: DNA  
 ORGANISM: *Staphylococcus epidermidis*  
 US-09-134-001C-2508

Query Match 6.0%; Score 37.2; DB 4; Length 576;  
 Best Local Similarity 44.7%; Pred. No. 1.2;  
 Matches 144; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

QY 210 GGTTCACAGACGTAGTATATGCTTTATTTATGTCGCAAGATGCTATTT 269  
 DB 87 GGTAAAGATCACAAAGACATGATTAATAAGTTGTTTATCATATTAAGTGC 146  
 QY 270 ATTGATGCCAATATTACTTTTATGCTGTTACGGCAGATTCATTACGTGCAAGT 329  
 DB 147 AATTATGACAGATTTTTCCTATGATGAGCAACCTTAATTTGACCCATGCTTT 206  
 QY 330 TGGTCATTATGACACTACCAAGGTTTAAATTAAATGAGTAATTAAATATATATA 389  
 DB 207 AGCGATGATAAGAAATTTAGTTCAATATAAAAGCTAGTTTCCCTTTATATAGCTTG 266  
 QY 390 AGGTTGAAGAAGATGTTGCTTCCCAACAACCTGTTGCACTTTACGATTTAGT 449  
 DB 267 TGGGTTATCGTATATTCACACATTTACCTGTTGTAAGTAATTTGTTAACTTGT 326  
 QY 450 TCAAGTAATTTGTTATAGATTTGTTCCATATATGATTAATAAGAGAGTTTCAACTT 509  
 DB 327 ATCAGCAGATTACTTTGTTGATTAATTTGTTGAATTTTAAAGCGATGTTTATACCTTA 386  
 QY 510 TTTACCATTAATATATGCAAGT 531  
 DB 387 TTTATTTCTCTATATAGCTAGT 408

## RESULT 38

US-08-213-419B-3/c  
 Sequence 3, Application US/08213419B  
 Patent No. 6333406  
 GENERAL INFORMATION:

APPLICANT: Inselburg, J. et al.  
 TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM  
 FILE REFERENCE: J11-002CNCIP  
 CURRENT APPLICATION NUMBER: US/08/213,419B  
 PRIOR FILING DATE: 1994-03-14  
 PRIOR APPLICATION NUMBER: US 07/870,506  
 PRIOR FILING DATE: 1992-04-17  
 NUMBER OF SEQ ID NOS: 20  
 SOFTWARE: Patent In Ver. 2.0  
 SEQ ID NO 3  
 LENGTH: 6124  
 TYPE: DNA  
 ORGANISM: *Plasmodium falciparum*  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (2407)..(2439)  
 NAME/KEY: CDS  
 LOCATION: (2598)..(3404)  
 NAME/KEY: CDS  
 LOCATION: (3580)..(3720)  
 NAME/KEY: CDS  
 LOCATION: (3850)..(5835)  
 US-08-213-419B-3

Query Match 6.0%; Score 37.2; DB 4; Length 6124;  
 Best Local Similarity 45.3%; Pred. No. 2.3;  
 Matches 135; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 144 AGCGTAATGAGACGCTATTGAAACAATTTTCTACTATATTTTACGAATCATTTGG 203

DB 2425 AGAAATATATGACTTCTATTTGATATATGAAATTAAGTATATAATTTTAAAGACT 2366  
 QY 204 GTTAGAGTTACAGACAGTATATGCTTTATTTATATGTCGCAAGATGAC 263  
 DB 2365 TTAATTTTGAATATGTTGATTTATTAATTTTCTTTATGTTATATATATATAT 2306  
 QY 264 TATTTATTTGATGCAATATTACTTTTATTTGCTGTTACGGCAGATTCATTACGTGT 323  
 DB 2305 TTAATATATATATATATATTTATTTTATTTTATTTTATTTTGGCAATCTAATTAATT 2246  
 QY 324 ACAAGTGTGCAATTATGACATCAAGGTTTAAATTAAATGAGATTAATTAAT 383  
 DB 2245 TAAATTAATAAATTTATATATATATATATATATATATATATATATATATAT 2186  
 QY 384 AATAAAGGTTGAAGAATGTTTCTTCCAAACAACCTGTTGACATTTACGT 441  
 DB 2185 TTTAAATTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 2128

## RESULT 39

US-09-599-360B-7  
 Sequence 7, Application US/09599360B  
 Patent No. 6548633  
 GENERAL INFORMATION:

APPLICANT: Dumas Milne Edwards, J.B.  
 APPLICANT: Bougueleret, L.  
 TITLE OF INVENTION: Complementary DNA's Encoding Proteins with Signal Peptides  
 FILE REFERENCE: GENSET.050CP3  
 CURRENT APPLICATION NUMBER: US/09/599,360B  
 PRIOR FILING DATE: 2000-06-21  
 PRIOR APPLICATION NUMBER: 60/113,686  
 PRIOR FILING DATE: 1998-12-22  
 PRIOR APPLICATION NUMBER: 60/141,032  
 PRIOR FILING DATE: 1999-06-25  
 PRIOR APPLICATION NUMBER: 09/469,099  
 PRIOR FILING DATE: 1999-12-21  
 NUMBER OF SEQ ID NOS: 123  
 SOFTWARE: Patent.pm  
 SEQ ID NO 7  
 LENGTH: 1918  
 TYPE: DNA  
 ORGANISM: *Homo sapiens*  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 238..612  
 NAME/KEY: sig peptide  
 LOCATION: 238..348  
 OTHER INFORMATION: Von Heijne matrix  
 OTHER INFORMATION: score 9.4  
 NAME/KEY: polyA signal  
 LOCATION: 1885..1890  
 NAME/KEY: polyA site  
 LOCATION: 1905..1918  
 NAME/KEY: misc feature  
 LOCATION: 945,1624  
 OTHER INFORMATION: n=a, g, c or t  
 US-09-599-360B-7

Query Match 5.9%; Score 37; DB 4; Length 1918;  
 Best Local Similarity 50.3%; Pred. No. 1.9;  
 Matches 91; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 338 TATGACTACAAAGGTTTAAATTAATGAGTAATTAATTAATTAATTAAGGTTGA 397  
 DB 1723 TTTGTAATATCGTGTCATTAATAATTAATTAATTAATTAATTAATTAATTAATTA 1782  
 QY 338 AAGAAATGTTGCTTCTCAACAACATTTGCACTTTTACGTATTTAGTTCAAGTAA 457  
 DB 1783 AATTATCTCTGATTTGTGAACATTAAGACTTTCTAGAAATTTACTTATTTCACTCTGATCTTA 1842

QY	458	TTGTTAATAGTATGTTCCATATATGATTAATTAAGGAGAGTTTCAACTTTTACCAT	517
Db	1843	AATTTTTCTAATGACACATATTACTTTGTATATGAGAAATTAATTAATGCTATATTTT	1902
QY	518	T 518	
Db	1903	T 1903	

RESULT 40

```

US-09-599-360B--60
: Sequence 60, Application US/09599360B
: Patent No. 6548633
: GENERAL INFORMATION:
: APPLICANT: Dumas Milne Edwards, J.B.
: APPLICANT: Bouguetleret, L.
: APPLICANT: Jobert, S.
: TITLE OF INVENTION: Complementary DNA's Encoding Proteins with Signal Peptides
: FILE REFERENCE: GENSET.050CP3
: CURRENT FILING DATE: 2000-06-21
: PRIOR APPLICATION NUMBER: 60/113,686
: PRIOR FILING DATE: 1998-12-22
: PRIOR APPLICATION NUMBER: 60/141,032
: PRIOR FILING DATE: 1999-06-25
: PRIOR APPLICATION NUMBER: 09/469,099
: PRIOR FILING DATE: 1999-12-21
: NUMBER OF SEQ ID NOS: 123
: SOFTWARE: Patent.pm
: SEQ ID NO 60
: LENGTH: 1918
: TYPE: DNA
: ORGANISM: Homo Sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 238..612
: NAME/KEY: sig_peptide
: LOCATION: 238..348
: OTHER INFORMATION: Von Heljne matrix
: OTHER INFORMATION: score 9.40
: OTHER INFORMATION: seq LLCVLSASQLS/QD
: NAME/KEY: polyA_signal
: LOCATION: 1885..1890
: NAME/KEY: polyA_site
: LOCATION: 1905..1918
: NAME/KEY: misc.feature
: LOCATION: 945..1624
: OTHER INFORMATION: n=a, g, c or t
: US-09-599-360B--60

```

Query Match	5.9%	Score 37;	DB 4;	Length 1918;	.
Best Local Similarity	50.3%;	Pred. No. 1.9;			
Matches 91; Conservative	0;	Mismatches 90;	Indels 0;	Gaps 0;	

QY	338	TATGACATCAAGAAGTTTAAATTTAAAGAGTAATTTAAATTAATAAAGGCTGA	397
Db	1723	TTTGTATACGTGTCATAAAAATTTAAATAGGGTAACAACATATTTTAAAGCAGTTA	1782
QY	398	AAGGATGTTTGCTTCCACAAACACTGTTGCACTTTTACGTAAGTTAGTCAAGTAA	457
Db	1763	AATTAATCTGCTATTTGTAAGACTTTCAGATTTTACTTAATCTGACTTA	1844
QY	458	TTGTTAATAGTATGTTCCATATATGATTAATTAAGAAGAGTTTCAAACTTTTACAT	517
Db	1843	AATTTTTCTAATGAACACATATATCTTTGTAATCAGAAAATATTAATGACATGATTTT	1902
QY	518	T 518	
Db	1903	T 1903	

Search completed: October 4, 2003, 05:14:34  
Job time : 61 secs

**THIS PAGE BLANK (USPTO)**